

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:51:11 ; Search time 3842.15 Seconds
(without alignments)
1807.663 Million cell updates/sec

Title: US-09-394-745-5950
Perfect score: 421
Sequence: 1 ggggtccaggcacgcgtccga.....agtggcagaatttggtgccgc 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*

28: em_un:*
 29: em_vi:*
 30: em_htgo_hum:*
 31: em_htgo_inv:*
 32: em_htgo_rod:*
 33: em_htg_hum:*
 34: em_htg_inv:*
 35: em_htg_rod:*
 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%		Query				Description
	No.	Score	Match	Length	DB	ID	
	1	87.6	20.8	1996	8	AF361600	AF361600 Arabidops
	2	78.8	18.7	104769	8	ATF9G14	AL162973 Arabidops
	3	78.8	18.7	107603	8	AC016661	AC016661 Arabidops
c	4	37.2	8.8	179681	9	AL365225	AL365225 Human DNA
	5	36.6	8.7	74881	2	AC020320	AC020320 Drosophil
c	6	36.6	8.7	114396	9	HS1098F8	AL034547 Human DNA
c	7	36.6	8.7	170869	3	AC011696	AC011696 Drosophil
c	8	36.6	8.7	171831	3	AC007473	AC007473 Drosophil
c	9	36.6	8.7	183215	2	AC023370	AC023370 Homo sapi
c	10	36.6	8.7	278196	3	AE003825	AE003825 Drosophil
	11	36.4	8.6	3911	5	SAU02975	U02975 Squalus aca
c	12	36.2	8.6	94555	9	AL589684	AL589684 Human DNA
c	13	36.2	8.6	168438	2	AC024632	AC024632 Homo sapi
	14	36	8.6	824	8	AF137266	AF137266 Nuphar lu
	15	35.6	8.5	101261	2	AC010628	AC010628 Homo sapi
	16	35.6	8.5	155185	2	AC023549	AC023549 Homo sapi
c	17	35.6	8.5	156886	2	AC023445	AC023445 Homo sapi
	18	35.6	8.5	168353	2	AL390024	AL390024 Homo sapi
c	19	35.4	8.4	103640	8	AP003278	AP003278 Oryza sat
c	20	35.4	8.4	154966	2	AP001387	AP001387 Homo sapi
c	21	35.4	8.4	155874	2	AC090415	AC090415 Homo sapi
	22	35.4	8.4	166350	2	AP003330	AP003330 Oryza sat
c	23	35.4	8.4	167348	2	AC027780	AC027780 Homo sapi
c	24	35.4	8.4	172693	2	AC074246	AC074246 Homo sapi
	25	35.4	8.4	176697	2	AC021170	AC021170 Homo sapi
c	26	35.4	8.4	192391	9	AC010768	AC010768 Homo sapi
c	27	34.8	8.3	1539	8	HTL2NFR	Z26251 H.tuberosus
c	28	34.8	8.3	95533	2	AC093224	AC093224 Homo sapi
c	29	34.8	8.3	162098	2	AC019032	AC019032 Homo sapi
c	30	34.8	8.3	162671	2	AC034299	AC034299 Homo sapi
	31	34.8	8.3	173334	2	AC091198	AC091198 Homo sapi
	32	34.8	8.3	173585	2	AC021113	AC021113 Homo sapi
c	33	34.8	8.3	216078	2	AC087053	AC087053 Homo sapi
c	34	34.6	8.2	145722	2	AC015503	AC015503 Homo sapi
c	35	34.6	8.2	162125	2	AC073494	AC073494 Homo sapi
c	36	34.6	8.2	165394	8	AC025296	AC025296 Oryza sat
c	37	34.6	8.2	166019	2	AC015648	AC015648 Homo sapi
	38	34.6	8.2	203203	9	AC020910	AC020910 Homo sapi

	39	34.4	8.2	172261	33	AC026978	Ac026978 Homo sapi
	40	34.4	8.2	176179	2	AC092591	AC092591 Homo sapi
	41	34.2	8.1	1452	1	UEU81722	U81722 Unidentifie
c	42	34.2	8.1	1753	9	AK023499	AK023499 Homo sapi
c	43	34.2	8.1	138322	2	AC012241	AC012241 Homo sapi
c	44	34.2	8.1	146072	2	AL360233	AL360233 Homo sapi
	45	34.2	8.1	161992	2	AC079750	AC079750 Homo sapi

ALIGNMENTS

```

RESULT      1
AF361600
LOCUS       AF361600      1996 bp      DNA              PLN      23-MAY-2001
DEFINITION  Arabidopsis thaliana AT5g02970/F9G14_280 gene, complete cds.
ACCESSION   AF361600
VERSION     AF361600.1  GI:13605548
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1996)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Banh,J., Bowser,L.,
            Carninci,P., Chung,M.K., Goldsmith,A.D., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M.,
            Lin,J., Liu,S.X., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Pham,P.K., Quach,H.L., Sakano,H., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Toriumi,M., Yamada,K., Yu,G., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL    Unpublished
REFERENCE   2  (bases 1 to 1996)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Banh,J., Bowser,L.,
            Carninci,P., Chung,M.K., Goldsmith,A.D., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M.,
            Lin,J., Liu,S.X., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Pham,P.K., Quach,H.L., Sakano,H., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Toriumi,M., Yamada,K., Yu,G., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL    Submitted (15-MAR-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C., Tracy,S.E., Banh,J.,
            Bowser,L., Chung,M.K., Goldsmith,A.D., Jones,T., Karlin-Neumann,G.,

```

Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
 Palm,C.J., Pham,P.K., Quach,H.L., Sakano,H., Southwick,A.,
 Tang,C.C., Toriumi,M., Yamada,K., Yu,G., Davis,R.W., Theologis,A.,
 and Ecker,J.R.

Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
 this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
 contributed equally to this work as PIs.

```

FEATURES             Location/Qualifiers
    source             1. .1996
                       /organism="Arabidopsis thaliana"
                       /db_xref="taxon:3702"
                       /chromosome="5"
                       /clone="RAFL09-10-O19(R12091)"
                       /note="ecotype: Columbia"
    5'UTR             1. .225
    CDS                226. .1770
                       /note="unknown protein"
                       /codon_start=1
                       /product="AT5g02970/F9G14_280"
                       /protein_id="AAK32768.1"
                       /db_xref="GI:13605549"
                       /translation="MEEIRGVPTWQEELASLV DAGLRYDGAPIDLTAATKRSGFVSAD
                       GSGSEPKETLKDQVTGFMKSWGEMLLELAGCKDIVQQT VVTDDSF LVRKLRKPAKV
                       SKKLSFLNEFLPEDRDPIHAWPVIFVFL LALAALSFSPENDRPVTVITKLR LHPTGA
                       TRVQLPDGRYIAYQELGVSAERARYSLVMPHSFLSSRLAGIPGVKKSL LVEYGVRLVS
                       YDLPGFGESDPHGRNLSSSASD MINLAAAIGIDEKFWLLGYSTGSIHTWAGMKYFPE
                       KIAGAAMVAPVINPYEPSMVKEEVVKTWEQWLTKRK FMYFLARRFPILLPFFYRRSFL
                       SGNLDQLDQWMALSLGEKDKLLIKDPTFQEVYQRNVEESVRQGITKPFVEEAVLQVSN
                       WGFTLSEFRTQKKCATNGVLSWLMSEAECE LIGFRKPIHIWQGMEDRVAPPSMSD
                       YISRMIP EATVHKIRNEG HFSFFYFCDECHRQIFYALFGEPK GQLERVKETEDTVVET
                       EAHKDT"
    3'UTR             1771. .1996
BASE COUNT           601 a      369 c      483 g      543 t
ORIGIN

```

Query Match 20.8%; Score 87.6; DB 8; Length 1996;
 Best Local Similarity 60.4%; Pred. No. 2.6e-16;
 Matches 180; Conservative 0; Mismatches 114; Indels 4; Gaps 2;

```

Qy 118 aggacagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgag 177
    |||| | | |||| | || ||| ||| | | | | | ||||| || | |
Db 1280 AGGATAAACTTTTAATCAAAGATCCAACGTTTCAAGAAGTTTATCAAAGGAACGTGGAGG 1339

Qy 178 agtctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatct 237
    | || | | | ||||| ||| ||||| || ||||| ||||| |||||
Db 1340 AATCAGT-CCGTCAAGGAATCACAAAACATTTGTAGAAGAAGCCGTGCTTCAAGTATCG 1398

Qy 238 gactggggtttcagcctatctgacatccaactgcagaagaaa---gaggctcaaggcttt 294
    | ||||| || | || || ||| ||| ||||| || | | | |
Db 1399 AATTGGGGCTTTACTCTTTCGGAATTCGCGACACAGAAGAAATGTGCAACCAACGGTGTC 1458

Qy 295 tttgaactcatcacgtctctgttcaatcatgctgaaaaacagtgggtgggatttctgggc 354
    || ||| ||| ||| || | | ||| || | |||||
Db 1459 CTTTCTTGCTCATGTCAATGTACAGTGAAGCCGAATGTGAACTAATCGGATTTCGAAAA 1518

```

Qy 355 ccaatacatatatcgaggggatagatgaccgagtgatctcgccctcagtggcagaat 412
 || || || |||| ||||| || || || ||||| | || ||| || || |
 Db 1519 CCCATTACATATGGCAGGGAATGGAGGATCGAGTGGCTCCACCATCAATGAGTGACT 1576

RESULT 2

ATF9G14

LOCUS ATF9G14 104769 bp DNA PLN 03-APR-2000

DEFINITION Arabidopsis thaliana DNA chromosome 5, BAC clone F9G14 (ESSA project).

ACCESSION AL162973

VERSION AL162973.1 GI:7413544

KEYWORDS .

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 104769)

AUTHORS Bevan,M., Terry,N., Ardiles,W., Buysshaert,C., Dasseville,R., De
 Clerck,R., De Keyser,A., Neyt,P., Rouze,P., Van Den Daele,H.,
 Villaroel,R., Gielen,J., Van Montagu,M., Bancroft,I., Mewes,H.W.,
 Rudd,S., Lemcke,K. and Mayer,K.F.X.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 104769)

AUTHORS EU Arabidopsis sequencing,project.

TITLE Direct Submission

JOURNAL Submitted (03-APR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk

COMMENT Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES Location/Qualifiers

source 1. .104769
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 misc_feature 1. .10970
 /note="overlap with BAC T22P11"
 gene complement(join(4857. .5159,5958. .6113,6244. .7155))
 /gene="F9G14_10"
 gene 4857. .7155
 /gene="F9G14_10"
 CDS complement(join(4857. .5159,5958. .6113,6244. .7155))
 /gene="F9G14_10"
 /note="similarity to various predicted proteins,
 Arabidopsis thaliana"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB86024.1"
 /db_xref="GI:7413545"
 /translation="MAKRRERRGRRQHRSHRRIQRIIDGADFINYPDDILHHILSFI"

```

PTDLAMRTSVLSRRWRHVWCETPCLDITLKHGAMNQTLTSYTAPIITSFKLVMDLNSN
TVPQVDSWIEFALSRNVQNLSVFVRDFTYSKTYRFPDIFYLSSSLKLLDVTLDFFDMI
PTCTVSWKSLRNLTLRFCQIPDESIHNILSGCPILES LTLDTCRLLERLDLSKSPNLR
RLDINQQYRRTPVAIVAPHIYYLRLTYSSPTSTIVDVSSLSEANLTI ISSLLSPLTA
DGYQTMALEMLSKFHNVKRLTVGETLLQILSLAELRGVPFPTLVKVTFTVRSV
IPGISRLLQNSPGLKKLRPSTMKMHHLKGLYPDQCWRSTCEVFPTSKEIYKMLGCNDA
TLKLVASFMDLVLRNAKTLERMVWVWLGGIYFNGDAPWFEEELFDMVETLSRNNVVSIL
LKQSNC"
exon      complement(4857. .5159)
          /gene="F9G14_10"
          /number=1
intron    complement(5160. .5957)
          /number=1
exon      complement(5958. .6113)
          /gene="F9G14_10"
          /number=2
intron    complement(6114. .6243)
          /number=2
exon      complement(6244. .7155)
          /gene="F9G14_10"
          /number=3
exon      8104. .8370
          /gene="F9G14_20"
          /number=1
gene      8104. .8956
          /gene="F9G14_20"
CDS       join(8104. .8370,8463. .8541,8629. .8716,8860. .8956)
          /gene="F9G14_20"
          /codon_start=1
          /product="putative protein"
          /protein_id="CAB86025.1"
          /db_xref="GI:7413546"
          /translation="MDMSLALSTAPMSRTIISATRRSQVSQPKAKKVKPANKRPTMST
          SGFSGGTTKELTWKCVGCGACCKIAKDFS FATPDEIFDNPDDVELYRSMIGDDGWCL
          NYDKATRKCSIYADRPYFCRVEPEVFKSLYGIEKKFNKEAVSCCIDTIKTIYGPDSK
          ELDSFNRAIRSNPSSS"
intron    8371. .8462
          /gene="F9G14_20"
          /number=1
exon      8463. .8541
          /gene="F9G14_20"
          /number=2
intron    8542. .8628
          /gene="F9G14_20"
          /number=2
exon      8629. .8716
          /gene="F9G14_20"
          /number=3
intron    8717. .8859
          /gene="F9G14_20"
          /number=3
exon      8860. .8956
          /gene="F9G14_20"
          /number=4
exon      complement(9258. .9481)
          /gene="F9G14_30"
          /number=1

```

```

gene      complement(join(9258. .9481,9597. .9789))
          /gene="F9G14_30"
gene      9258. .9789
          /gene="F9G14_30"
CDS       complement(join(9258. .9481,9597. .9789))
          /gene="F9G14_30"
          /note="similarity to predicted protein, Arabidopsis
          thaliana"
          /codon_start=1
          /product="putative protein"
          /protein_id="CAB86026.1"
          /db_xref="GI:7413547"
          /translation="MTGIGGEKCKRCGIYEQGS�VSDKEFDVWEVCPTDFSASQVYMH
          FKEKEINATFVCHGCAKFHSAVAASSPQEEGYNGLTFMIAIIAGVLCTTLVVVGGVFM
          FKHTQRMKKQRDQARFMQLFEESDEPEDELGLDPVI"
intron    complement(9482. .9596)
          /number=1
exon      complement(9597. .9789)
          /gene="F9G14_30"
          /number=2
tRNA      10120. .10201
          /note="tRNA predict as a tRNA- Ser : anticodon cga"
CDS       11086. .11703
          /gene="F9G14_40"
          /note="similarity to pathogenesis-related protein 1.2,
          Triticum aestivum, EMBL:AJ007349)"
          /codon_start=1
          /product="pathogenesis related protein-like"
          /protein_id="CAB86027.1"
          /db_xref="GI:7413548"
          /translation="MSSSSLYHPFCLSISSVLLLLLLIFSGEFPSTAGTSSPDTKAAA
          ARATNRGRNRNKQSAEFLAHNAARVASGASNLRWQGLARFASKWAKQRKSDCKMTHS
          GGPYGENIFRYQRSENWSPRRVVDKWMDESLNYDRVANTCKSGAMCGHYTQIVWRTTT
          AVGCARSKCDNNRGFLVICEYSPSGNYEGESPFDPKITLKLDKI"
exon      11086. .11703
          /gene="F9G14_40"
          /number=1
gene      11086. .11703
          /gene="F9G14_40"
gene      11993. .13743
          /gene="F9G14_50"
exon      11993. .12170
          /gene="F9G14_50"
          /number=1
CDS       join(11993. .12170,12466. .12607,12653. .12806,12920.
.12967,
          13093. .13193,13611. .13743)
          /gene="F9G14_50"
          /codon_start=1
          /product="putative protein"
          /protein_id="CAB86028.1"
          /db_xref="GI:7413549"
          /translation="MSITRSAIRSSLCLRKLDPEISRSSLPFLQEWKCLSTATEQPP
          PASPLPPPPGGSPGEGRFYKGKFSKHALKTDIMNVLEGCSLTSDDLKFNYPGGNL
          TPAAVEEFESKLMNFYPFSFSTMLVYISFVQFPSSLAYDKALRNIAKKGKLYRLEKAA
          RAQWDPIVPYEGKVVALHGIPVNAITDDIDRFLSGCLYYPGSIQFLTQVQLGTSKRVA
          LVRFTSQTQAMNAYITKNRNFLNQRITLQVLQ"

```

intron	12171. .12465
	/gene="F9G14_50"
	/number=1
exon	12466. .12607
	/gene="F9G14_50"
	/number=2
intron	12608. .12652
	/gene="F9G14_50"
	/number=2
exon	12653. .12806
	/gene="F9G14_50"
	/number=3
intron	12807. .12919
	/gene="F9G14_50"
	/number=3
exon	12920. .12967
	/gene="F9G14_50"
	/number=4
intron	12968. .13092
	/gene="F9G14_50"
	/number=4
exon	13093. .13193
	/gene="F9G14_50"
	/number=5
intron	13194. .13610
	/gene="F9G14_50"
	/number=5
exon	13611. .13743
	/gene="F9G14_50"
	/number=6
exon	15604. .16428
	/gene="F9G14_60"
	/number=1
gene	15604. .16428
	/gene="F9G14_60"
CDS	15604. .16428
	/gene="F9G14_60"

Query Match 18.7%; Score 78.8; DB 8; Length 104769;
 Best Local Similarity 60.7%; Pred. No. 2.1e-13;
 Matches 164; Conservative 0; Mismatches 102; Indels 4; Gaps 2;

Qy	110	cattattcaggacagaacttttactggaacgtcctgtgttcaatgcattctgggaaaggaa	169
Db	92312	CATCTTTCAGGATAAACTTTTAATCAAAGATCCAACGTTTCAAGAAGTTTATCAAAGGAA	92371
Qy	170	tgttgcagagtctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgc	229
Db	92372	CGTGGAGGAATCAGT-CCGTCAAGGAATCACAAAACCATTTGTAGAAGAAGCCGTGCTTC	92430
Qy	230	aagtatctgactgggggtttcagcctatctgacatccaactgcagaagaaa---gaggctc	286
Db	92431	AAGTATCGAATTGGGGCTTTACTCTTTCGGAATTCGCACACAGAAGAAATGTGCAACCA	92490
Qy	287	aaggcttttttgaactcatcacgtctctgttcaatcatgctgaaaaacagtgggtgggat	346
Db	92491	ACGGTGTCTCTTCTTGGCTCATGTCAATGTACAGTGAAGCCGAATGTGAACTAATCGGAT	92550

Qy 347 ttctggggcccaatacatatatcgagggga 376
||| || || || |||| |||| ||
Db 92551 TTCGAAACCCATTACATATGGCAGGTGA 92580

RESULT 3

AC016661

LOCUS AC016661 107603 bp DNA PLN 16-MAR-2001

DEFINITION Arabidopsis thaliana chromosome 3 BAC F11F8 genomic sequence,
complete sequence.

ACCESSION AC016661

VERSION AC016661.7 GI:12484383

KEYWORDS HTG.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 107603)

AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.

TITLE Arabidopsis thaliana chromosome 3 BAC F11F8 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 107603)

AUTHORS Lin,X. and Kaul,S.

TITLE Direct Submission

JOURNAL Submitted (04-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org

REFERENCE 3 (bases 1 to 107603)

AUTHORS Lin,X.

TITLE Direct Submission

JOURNAL Submitted (25-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

REFERENCE 4 (bases 1 to 107603)

AUTHORS Town,C.D. and Kaul,S.

TITLE Direct Submission

JOURNAL Submitted (16-MAR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

COMMENT On Jan 25, 2001 this sequence version replaced gi:12280748.

Address all correspondence to:at@tigr.org

BAC clone F11F8 is from Arabidopsis thaliana chromosome 3
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.

Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm_hm/glimmerm.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with

similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

```

FEATURES             Location/Qualifiers
    source            1. .107603
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /map="m532"
                     /clone="F11F8"
    misc_feature      1. .8479
                     /note="overlap with BAC clone F3L24
                     (AC011436:98210. .106688)."
```

mRNA	complement(join(<1631. .3543,3793. .>4006))
gene	complement(1631. .4006)
CDS	complement(join(1808. .3543,3793. .4006))

```

                     /gene="F11F8.1"
                     /note="identical to GB:CAA76606"
                     complement(join(1808. .3543,3793. .4006))
                     /gene="F11F8.1"
                     /codon_start=1
                     /product="heat shock cognate 70kD protein"
                     /protein_id="AAF23276.1"
                     /db_xref="GI:6682224"
                     /translation="MAGKGEGPAIGIDLGTTYSCVGVWQHDRVEIIANDQGNRTTPSY
                     VAFTDSERLIGDAAKNQVAMNPINTVFDKRLIGRRFTDSSVQSDIKLWPFTLKSGPA
                     EKPMIVVNYKGEDKEFSAEEISSMILIKMREIAEAYLGTTIKNAVVTVPAYFNDSQRQ
                     ATKDAGVIAGLNMRIINEPTAAAIAYGLDKKATSVGEKNVLI FDLGGGTFDVSLLTI
                     EEGI FEVKATAGDTHLGGEDFDNRMVNH FVQEFKRKNKKDISGNPRALRRLRTACERA
                     KRTLSSTAQTTIEIDSLFDGIDFYAPITRARFEELNIDLFRKCEPVEKCLRDAMDK
                     NSIDDVVLVGGSTRIPKVQQLLVDFNGKELCKSINPDEAVAYGA AVQAAILSGEGNE
                     KVQDLLLLDVTPLSLGLTAGGVMTVLIQRNTTIPTKKEQVFSTYSDNQPGVLIQVYE
                     GERARTKDNLLGKFELSGIPPAPRGVPQITVCFDIDANGILNVSAEDKTTGQKNKIT
                     ITNDKGRLSKDEIEKMQEAEKYKSEDEEHKKKVDAKNALENYAYNMRNTIRDEKIGE
                     KLAGDDKKKIEDSIEAAIEWLEANQLAECDEFEDKMKELESICNP IIAKMYQGGEAGG
                     PAAGGMDEDVPPSAGGAGPKIEEVD"
    mRNA              join(<5922. .6455,6539. .6706,6840. .>8465)
                     /gene="F11F8_2"
    gene              5922. .8465
                     /gene="F11F8_2"
                     /note="predicted by genscan , similar to hypothetical
                     protein GB:AAF14039"
    CDS               join(5922. .6455,6539. .6706,6840. .8465)
                     /gene="F11F8_2"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAF23277.1"
                     /db_xref="GI:6682225"
                     /translation="MRDTTWLERLGLALRTAMACLIVSLTTLTGPKPLRHFTTFFAFS
                     YLTTILIWLSDAEPTYGEVLKCCLDVSYATFQTIAIALVSVLVVGPASLGNGLVAPVA"

```

```

VALASFIVAFPVSTSLLTAKRIAAGQIVVVYVTFVVFNGEVAHVFMPLPVHVAGSTALGA
IASLIAVLLPFPRLAHSQMSKGCKLYAENALERLNMFVEIMMARDNTTAQVLIARAAS
LSAAAKNTLKNIKIHHERISWERPDTRFLSRKQKLDPAEKLHATDFLLRGLELALGSC
SSFPQGMSRDELTRLLEGPRTHIAPRSESTLKSQDSLQGWHEAESLSTAALPVCFFRY
CVELFRGDFLSLRQDSKSVNGRTTEEEIHPANEGLSMARKFWDILCVWMARERFVFAF
KCSISLGLAVLFGILYNKNNGYWSGLTVAISLVSGRQATLTVANSRLQGTAMGSVYGL
ICCSVFQRLEEFRLPLLPWIIILAVFMRHSKVYGPQGGVTAIAIALLILGRNYGAPT
EFAIARIVEASIGLLCFVFGIILVT PARAATLARTEISHCLDALLDCIQSLVLCSEQK
NQKVVDLRKSQVKLKSHVEALERFAAEALTEPKIPFLRLNTDSYNRLGFSFSKISD
LCLYVCDGLKNLSGVQPTLAFPWDNITHELRAFQEKLHPSVKCLKEISQTKSQARLQK
ELQKRKICHDEAGTTSNDNYSYMELGPSQADVERFSVSFVMLLKEATDKISCNTADD
AFKSETALCLSSLGFCISRLMQETICIMTEITHTT"

mRNA      complement(<9001. .>9108)
          /gene="F11F8_3"
gene      complement(9001. .9108)
          /gene="F11F8_3"
CDS       complement(9001. .9108)
          /gene="F11F8_3"
          /codon_start=1
          /product="unknown protein"
          /protein_id="AAF23278.1"
          /db_xref="GI:6682226"
          /translation="MFDDQDLGFFANFLGIFIFILVIAYHFVMADPKFE"

mRNA      complement(join(<10934. .11033,11335. .11585,11949. .12062,
12157. .12220,12325. .12482,12614. .>12784))
          /gene="F11F8_4"
gene      complement(10934. .12784)
          /gene="F11F8_4"
          /note="predicted by genefinder"
CDS       complement(join(10934. .11033,11335. .11585,11949. .12062,
12157. .12220,12325. .12482,12614. .12784))
          /gene="F11F8_4"
          /codon_start=1
          /product="hypothetical protein"
          /protein_id="AAF23279.1"
          /db_xref="GI:6682227"
          /translation="MESRNDEEAPLISASGEDRKVRAGKCYTRDVHILSISFLLIFLA
YGAAQNLETTVNKDLGTISLGILYVSFMFCMSVASLVVRLMGSKNALVLGTTGYWLFV
AANLKPSWFTMPASLYLGFAASIIWVGQGTYLTSIARSHATDHGLHEGSGVIGVFNGE
FWAMFACHQEGSTSGTTLMLVFLFSMTLGTILMFFIRKIDGEDGKGPGVSPVGLVDS
LASLPRMIITPLDIRMLLIVPLLAYSGLQQAQFVWAEFTKEIVTPAIGVSGVGAMAV
YGALDAVVS"

mRNA      complement(<13264. .>13644)
          /gene="F11F8_5"
gene      complement(13264. .13644)
          /gene="F11F8_5"
          /note="similar to histone H2B 3 GB:CAA12231 from
(Lycopersicon esculentum)"
CDS       complement(13264. .13644)
          /gene="F11F8_5"
          /codon_start=1
          /product="putative histone H2B"
          /protein_id="AAF23280.1"
          /db_xref="GI:6682228"
          /translation="MAPKAEKKPSEKAPKADKKITKEGGSERKKKTKKSTETYKIYLF
KVLKQVHPDIGISGKAMGIMNSFINDTFEKIALESSRLARYNKKPTITSREIQTAVRL
VLPGELAKHAVSEGTKAVTKFTSS"

```


Db 72535 GTCCTCTCTTGGCTAATGTCAATGTACAGTGAATCCGAATGTGAACTAATTGGTTTTTCGG 72594

Qy 352 ggccaatacatatatcgagg 373

|| ||||| |||| |||||

Db 72595 AAACCTATACACATATGGCAGG 72616

RESULT 4

AL365225/c

LOCUS AL365225 179681 bp DNA PRI 06-JUN-2001

DEFINITION Human DNA sequence from clone RP11-179A5 on chromosome 1, complete sequence.

ACCESSION AL365225

VERSION AL365225.12 GI:14329978

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 179681)

AUTHORS Hall, R.

TITLE Direct Submission

JOURNAL Submitted (03-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Jun 8, 2001 this sequence version replaced gi:14280436. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-179A5 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6 This sequence is the entire insert of clone RP11-179A5 The true left end of clone RP11-512F24 is at 131246 in this sequence. The true right end of clone RP4-658C17 is at 82886 in this sequence.

FEATURES Location/Qualifiers

source 1. .179681

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-179A5"
/clone_lib="RPCI-11.1"
repeat_region 1378. .1685
/note="HAL1 repeat: matches 1021. .1356 of consensus"
repeat_region 1695. .1828
/note="L1MD3 repeat: matches 7590. .7738 of consensus"
repeat_region 1836. .2066
/note="L1MD2 repeat: matches 5889. .6122 of consensus"
repeat_region 2067. .2357
/note="AluSp repeat: matches 1. .293 of consensus"
repeat_region 2358. .2453
/note="L1MD2 repeat: matches 5795. .5890 of consensus"
repeat_region 2534. .2708
/note="MIR repeat: matches 72. .261 of consensus"
repeat_region 3700. .3861
/note="MIR repeat: matches 44. .212 of consensus"
repeat_region 3864. .4681
/note="L1MC5 repeat: matches 7056. .7913 of consensus"
repeat_region 4697. .4849
/note="L2 repeat: matches 1376. .1526 of consensus"
repeat_region 5364. .5449
/note="43 copies 2 mer tc 62% conserved"
repeat_region 6192. .6269
/note="3 copies 26 mer 80% conserved"
repeat_region 6208. .6267
/note="30 copies 2 mer ca 90% conserved"
repeat_region 8988. .9265
/note="AluSc repeat: matches 1. .281 of consensus"
repeat_region 10039. .10232
/note="L1PA7 repeat: matches 5951. .6143 of consensus"
repeat_region 10509. .10634
/note="L1PA7 repeat: matches 6018. .6143 of consensus"
repeat_region 10929. .11087
/note="MLT1G repeat: matches 31. .179 of consensus"
repeat_region 11465. .11644
/note="MER20 repeat: matches 1. .217 of consensus"
repeat_region 13522. .13877
/note="MLT2FB repeat: matches 2. .366 of consensus"
repeat_region 14273. .14667
/note="MSTD repeat: matches 1. .394 of consensus"
repeat_region 14884. .15088
/note="L2 repeat: matches 2495. .2710 of consensus"
repeat_region 17250. .17341
/note="23 copies 4 mer tcct 80% conserved"
repeat_region 17753. .17907
/note="MIR repeat: matches 17. .192 of consensus"
repeat_region 18170. .18505
/note="MER2 repeat: matches 1. .345 of consensus"
repeat_region 18903. .19195
/note="L1MC3 repeat: matches 7357. .7735 of consensus"
repeat_region 19838. .20022
/note="AluY repeat: matches 124. .305 of consensus"
repeat_region 20309. .20472
/note="L1M4 repeat: matches 3852. .4017 of consensus"

```

repeat_region	20495. .21223 /note="L2 repeat: matches 1363. .2199 of consensus"
repeat_region	22711. .22834 /note="MER5B repeat: matches 54. .176 of consensus"
repeat_region	22835. .22874 /note="20 copies 2 mer gt 100% conserved"
repeat_region	22889. .23170 /note="AluJb repeat: matches 1. .282 of consensus"
repeat_region	25201. .25252 /note="Alu repeat: matches 251. .302 of consensus"
repeat_region	25257. .26124 /note="L1MC5 repeat: matches 6954. .7777 of consensus"
repeat_region	26136. .26611 /note="L1MC/D repeat: matches 5316. .6983 of consensus"
repeat_region	26852. .28167 /note="L1MC3 repeat: matches 6435. .7739 of consensus"
repeat_region	28752. .28861 /note="L1MC/D repeat: matches 5484. .5583 of consensus"
repeat_region	29223. .29533 /note="AluY repeat: matches 1. .311 of consensus"
repeat_region	29534. .29654 /note="L2 repeat: matches 2378. .2491 of consensus"
repeat_region	29687. .29836 /note="MIR repeat: matches 31. .218 of consensus"
repeat_region	29883. .30020 /note="L2 repeat: matches 2162. .2295 of consensus"
repeat_region	30211. .30508 /note="AluSx repeat: matches 1. .298 of consensus"
repeat_region	31890. .32233 /note="L2 repeat: matches 1804. .2145 of consensus"
repeat_region	32234. .32544 /note="AluSq repeat: matches 1. .311 of consensus"
repeat_region	32545. .33119 /note="L2 repeat: matches 2145. .2750 of consensus"
repeat_region	34189. .34352 /note="L2 repeat: matches 691. .851 of consensus"
repeat_region	34606. .34936 /note="L2 repeat: matches 2130. .2481 of consensus"
repeat_region	34937. .35158 /note="MER45C repeat: matches 1. .255 of consensus"
repeat_region	35273. .35461 /note="MER45C repeat: matches 741. .952 of consensus"
repeat_region	35462. .35704 /note="L2 repeat: matches 2474. .2750 of consensus"
repeat_region	36266. .36356 /note="L2 repeat: matches 2409. .2502 of consensus"
repeat_region	36501. .36560 /note="L2 repeat: matches 2646. .2705 of consensus"
repeat_region	37841. .38131 /note="AluSp repeat: matches 1. .297 of consensus"
repeat_region	38207. .38468 /note="AluJb repeat: matches 1. .301 of consensus"
misc_feature	complement(38675. .39177) /note="match: STS: Em:HSPE17E05"
repeat_region	39185. .39486 /note="AluSg repeat: matches 1. .302 of consensus"
repeat_region	39632. .40051

```

repeat_region    /note="L1PB1 repeat: matches 5710. .6146 of consensus"
                  40063. .40380
repeat_region    /note="AluJb repeat: matches 1. .295 of consensus"
                  40388. .40527
repeat_region    /note="L1MC5 repeat: matches 7654. .7779 of consensus"
                  40659. .41529
repeat_region    /note="L1MD2 repeat: matches 5430. .6329 of consensus"
                  41710. .41835
repeat_region    /note="L1MB8 repeat: matches 5833. .5955 of consensus"
                  41836. .42139
repeat_region    /note="AluY repeat: matches 1. .302 of consensus"
                  42140. .42696
repeat_region    /note="L1MB8 repeat: matches 5258. .5833 of consensus"
                  42783. .43143
repeat_region    /note="L2 repeat: matches 2346. .2710 of consensus"
                  45083. .45209
repeat_region    /note="MIR repeat: matches 71. .207 of consensus"
                  45441. .45486
repeat_region    /note="23 copies 2 mer tt 73% conserved"
                  45488. .45705
repeat_region    /note="AluJo repeat: matches 39. .256 of consensus"
                  47081. .47215
repeat_region    /note="MIR repeat: matches 47. .191 of consensus"
                  47260. .47496
repeat_region    /note="MIR repeat: matches 1. .262 of consensus"
                  48297. .49263
repeat_region    /note="L2 repeat: matches 1727. .2701 of consensus"
                  49271. .49322
repeat_region    /note="L2 repeat: matches 2649. .2700 of consensus"
                  52241. .52671
repeat_region    /note="MLT1H repeat: matches 73. .547 of consensus"
                  52971. .53230
repeat_region    /note="AluJb repeat: matches 10. .289 of consensus"
                  53255. .53564
repeat_region    /note="AluSq repeat: matches 2. .311 of consensus"
                  53749. .53986
repeat_region    /note="MIR repeat: matches 3. .254 of consensus"
                  55654. .55770

```

```

Query Match      8.8%; Score 37.2; DB 9; Length 179681;
Best Local Similarity 47.4%; Pred. No. 2.1;
Matches 111; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```

```

Qy      10  cacgcgtccgaattgaggttagcttaacaattcttagtagtcaccccttcgattaaatgt 69
          || ||  | || | | || ||  |||| | || | | || | | |||| ||
Db 137319 CAGGCAATCAAAATAAAGTGAGTCACACAAATGTTTTTGTTCCTCCAGTACATCTAAAAGT 137260

Qy      70  caacatttgccttttcgcgttccaattactaatgttacggcattattcaggacagaactt 129
          |  ||  || |||  | ||||  || | ||||| | ||||
Db 137259 TATGTTTACACTATTCTTTAGTCTATTAAGTGTGCAATAGCATTATGTATTTAAGAACAA 137200

Qy     130  tactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgagagtcgtgtgcagc 189
          ||  || | | | | || || | || | || | | || | |
Db 137199 TATACATACATTATTTTTTAAATACTTTATTGCTAAAATATGCTAACAAACATCTGAGCC 137140

Qy     190  caaggagatgcaaggccatttgtggacgaagctgtgtgctgcaagtatctgactgg 243
          | ||  ||  | |||  || |||  || || ||||

```


Db 137139 TTCAGTGAGTCATAATCTTTTGCTGGTGGAGGGTCTTGCCTTGATGTGGCTGG 137086

Query Match 8.7%; Score 36.6; DB 2; Length 74881;
Best Local Similarity 51.5%; Pred. No. 3;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

RESULT	6			
HS1098F8/c				
LOCUS	HS1098F8	114396 bp	DNA	PRI 15-MAR-2001
DEFINITION	Human DNA sequence from clone RP5-1098F8 on chromosome 20p11.23-12.3. Contains an STS and GSSs, complete sequence.			

ACCESSION AL034547
 VERSION AL034547.14 GI:11139873
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 114396)
 AUTHORS Barlow, K.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
 requests: clonerequest@sanger.ac.uk

COMMENT On Nov 13, 2000 this sequence version replaced gi:9795173.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
 IMPORTANT: This sequence is not the entire insert of clone
 RP5-1098F8 It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP5-1098F8 is at 114396 in this
 sequence. The true right end of clone RP4-742J24 is at 100 in this
 sequence. This sequence was finished as follows unless otherwise
 noted: all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.
 RP5-1098F8 is from the library RPCI-5 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2.

FEATURES Location/Qualifiers
 source 1..114396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="p11.23-12.3"
 /clone="RP5-1098F8"
 /clone_lib="RPCI-5"
 repeat_region 371..461
 /note="L2 repeat: matches 2627..2729 of consensus"
 repeat_region 819..885
 /note="MIR repeat: matches 79..146 of consensus"

repeat_region	2210. .2339 /note="L2 repeat: matches 2188. .2339 of consensus"
repeat_region	2340. .2792 /note="MLT1F repeat: matches 44. .512 of consensus"
repeat_region	3012. .3148 /note="L2 repeat: matches 2553. .2656 of consensus"
repeat_region	3149. .3442 /note="AluSq repeat: matches 1. .296 of consensus"
repeat_region	3443. .3527 /note="L2 repeat: matches 2656. .2749 of consensus"
misc_feature	complement(5173. .5567) /note="match: GSS: Em:AQ136848"
repeat_region	6571. .6635 /note="MLT1F repeat: matches 14. .83 of consensus"
repeat_region	6884. .6907 /note="12 copies 2 mer aa 100% conserved"
repeat_region	7657. .8052 /note="MLT1H repeat: matches 2. .426 of consensus"
repeat_region	8740. .9303 /note="MLT2D repeat: matches 4. .552 of consensus"
repeat_region	9530. .9580 /note="MLT1I repeat: matches 96. .138 of consensus"
repeat_region	9612. .9825 /note="L2 repeat: matches 1655. .1871 of consensus"
repeat_region	10324. .10667 /note="L2 repeat: matches 1128. .1491 of consensus"
repeat_region	10841. .11012 /note="MER5B repeat: matches 1. .178 of consensus"
repeat_region	11174. .11277 /note="MER5B repeat: matches 74. .177 of consensus"
repeat_region	11278. .11322 /note="MLT1C repeat: matches 1. .45 of consensus"
repeat_region	11323. .11601 /note="AluSg repeat: matches 1. .308 of consensus"
repeat_region	11602. .12030 /note="MLT1C repeat: matches 45. .466 of consensus"
repeat_region	12031. .12095 /note="MER5B repeat: matches 10. .74 of consensus"
repeat_region	12793. .12881 /note="MER5A repeat: matches 9. .94 of consensus"
repeat_region	12882. .13252 /note="MLT1B repeat: matches 3. .390 of consensus"
repeat_region	13253. .13271 /note="MER5A repeat: matches 94. .112 of consensus"
repeat_region	13340. .13515 /note="MER5A repeat: matches 15. .186 of consensus"
repeat_region	13530. .13661 /note="MER5A repeat: matches 26. .184 of consensus"
repeat_region	13813. .15648 /note="MER52A repeat: matches 202. .1755 of consensus"
repeat_region	15673. .15837 /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region	16288. .16388 /note="MLT1J repeat: matches 407. .512 of consensus"
repeat_region	16428. .16716 /note="AluSx repeat: matches 1. .289 of consensus"
misc_feature	complement(17260. .17705)

repeat_region /note="match: GSS: Em:AQ764260"
19143. .19198
/note="28 copies 2 mer ta 69% conserved"
repeat_region 19205. .19284
/note="40 copies 2 mer at 70% conserved"
repeat_region 19207. .19286
/note="20 copies 4 mer atat 70% conserved"
repeat_region 19863. .20017
/note="MER5A repeat: matches 4. .156 of consensus"
repeat_region 20024. .20231
/note="MER5A repeat: matches 2. .187 of consensus"
repeat_region 20769. .21078
/note="AluYa5 repeat: matches 1. .310 of consensus"
repeat_region 21366. .21820
/note="MLT1C repeat: matches 1. .462 of consensus"
repeat_region 21999. .22254
/note="LTR16C repeat: matches 48. .318 of consensus"
repeat_region 22925. .23011
/note="L1MB6 repeat: matches 6079. .6164 of consensus"
repeat_region 23012. .23342
/note="AluJo repeat: matches 1. .312 of consensus"
repeat_region 23343. .23851
/note="L1MB6 repeat: matches 5550. .6079 of consensus"
repeat_region 24000. .24092
/note="MER69 repeat: matches 1. .105 of consensus"
misc_feature 25592. .26017
/note="match: GSS: Em:B44047"
misc_feature 25592. .26084
/note="match: GSS: Em:AQ055943"
repeat_region 26402. .26641
/note="MLT1H repeat: matches 1. .266 of consensus"
repeat_region 26642. .26946
/note="L2 repeat: matches 2453. .2750 of consensus"
misc_feature complement(28404. .28755)
/note="match: GSS: Em:AQ170827"
misc_feature 28774. .29229
/note="match: GSS: Em:AQ424970"
repeat_region 29256. .29368
/note="LTR16C repeat: matches 260. .372 of consensus"
repeat_region 29804. .29847
/note="11 copies 4 mer acac 84% conserved"
repeat_region 29812. .29847
/note="18 copies 2 mer ac 91% conserved"
repeat_region 30731. .30794
/note="16 copies 4 mer ctat 95% conserved"
repeat_region 31199. .31720
/note="L1ME1 repeat: matches 5549. .6099 of consensus"
repeat_region 32580. .32762
/note="MER5A repeat: matches 1. .189 of consensus"
misc_feature complement(33816. .34227)
/note="match: GSS: Em:AQ126427"
repeat_region 34058. .34566
/note="MLT1D repeat: matches 1. .505 of consensus"
misc_feature 34228. .34906
/note="match: GSS: Em:AQ343317"
misc_feature 34228. .34630
/note="match: GSS: Em:AQ122515"

Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
 Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
 Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
 Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
 Phouanenvong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
 Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

TITLE Sequencing of Drosophila chromosome 2R, region 48A-48C
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 170869)
 AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.

TITLE Direct Submission
 JOURNAL Submitted (11-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 COMMENT On Mar 23, 2001 this sequence version replaced gi:6119492.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgp@fruitfly.berkeley.edu.

FEATURES Location/Qualifiers
 source 1. .170869
 /organism="Drosophila melanogaster"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="48A-48C"
 /clone="BACR35F01 (D1156)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACe3.6)"

BASE COUNT 49941 a 35331 c 35461 g 50136 t
 ORIGIN

Query Match 8.7%; Score 36.6; DB 3; Length 170869;
 Best Local Similarity 51.5%; Pred. No. 3.2;
 Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 143 tgtgtttcaatgcattctgtgggaaaggaatgttgcagagtctgtgcagccaaggagatgcaa 202
 ||| || || | | ||||| ||| ||| | | | ||| ||| | | ||

Db 45086 TGTCTTGAAGTCTTCCTGGGAAATGAAAATTGAATAAGACATACAGACAAAAATACAAA 45027

Qy 203 ggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttcagcctatctgaca 262
| | | | | | | | | | | | | | | | | | | | | |

Db 45026 TTAATGCTATTTCAGGCAGCTGTTTGCATCGATTCCGAATAAAGTTTTTACCAATTTAACT 44967

Qy 263 tccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
| | | | | | | | | | | | | | | | | | | | | |

Db 44966 TTAAATGTAATAAAAAATAACTCAAATACTAATAAAGCTTAT 44924

RESULT 8
AC007473/c
LOCUS AC007473 171831 bp DNA INV 28-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 2R, region 48A-48B, BAC clone BACR38D12, complete sequence.
ACCESSION AC007473
VERSION AC007473.10 GI:13162476
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 171831)
AUTHORS Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferriera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
TITLE Sequencing of Drosophila chromosome 2R, region 48A-48B
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171831)
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Feb 28, 2001 this sequence version replaced gi:5670618.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu.

FEATURES Location/Qualifiers
 source 1. .171831
 /organism="Drosophila melanogaster"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="48A-48B"
 /clone="BACR38D12 (D590)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACe3.6)"
BASE COUNT 49756 a 35946 c 36141 g 49988 t
ORIGIN

Query Match 8.7%; Score 36.6; DB 3; Length 171831;
Best Local Similarity 51.5%; Pred. No. 3.2;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 143 tgtgttcaatgcattctgggaaggaatgttgagagctctgtgcagccaaggagatgcaa 202
 ||| || || | | ||||| ||| ||| | | | ||| ||| | | ||
Db 85405 TGTCTTGAAGTCTTCCTGGGAAATGAAAATTGAATAAGACATACAGACAAAAATACAAAA 85346

Qy 203 ggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttcagcctatctgaca 262
 | | | ||||| | | ||| | ||| | ||| | |||
Db 85345 TTAATGCTATTTCAGGCAGCTGTTTGCATCGATTCCGAATAAAGTTTTTACCAATTTAACT 85286

Qy 263 tccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
 | || | || || |||| | | | |||
Db 85285 TTAAATGTAATAAAAAATAACTCAAATACTAATAAAGCTTAT 85243

RESULT 9
AC023370/c
LOCUS AC023370 183215 bp DNA HTG 15-FEB-2001
DEFINITION Homo sapiens clone RP11-20L19, WORKING DRAFT SEQUENCE, 11 unordered
 pieces.
ACCESSION AC023370
VERSION AC023370.4 GI:12831386
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183215)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens, clone RP11-20L19
 JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183215)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavkiy, L.,
 Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
 Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Landers, T., Largocque, K., Lehoczký, J., Levine, R.,
 Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
 McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Olivar, T.M.,
 Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Feb 15, 2001 this sequence version replaced gi:10280847.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L3926
 Center clone name: 20_L_19
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 47% of reads
 Sequencing vector: Plasmid; n/a; 53% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 173128 bases at least Q40
 Consensus quality: 177123 bases at least Q30
 Consensus quality: 179772 bases at least Q20
 Insert size: 170000; agarose-fp
 Insert size: 182215; sum-of-contigs
 Quality coverage: 6.8 in Q20 bases; agarose-fp
 Quality coverage: 6.3 in Q20 ba.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 2147: contig of 2147 bp in length
 * 2148 2247: gap of 100 bp

```

*      2248      3360: contig of 1113 bp in length
*      3361 3460: gap of      100 bp
*      3461      4570: contig of 1110 bp in length
*      4571 4670: gap of      100 bp
*      4671      5920: contig of 1250 bp in length
*      5921 6020: gap of      100 bp
*      6021      7417: contig of 1397 bp in length
*      7418 7517: gap of      100 bp
*      7518      8558: contig of 1041 bp in length
*      8559 8658: gap of      100 bp
*      8659      9869: contig of 1211 bp in length
*      9870 9969: gap of      100 bp
*      9970     12680: contig of 2711 bp in length
*     12681 12780: gap of      100 bp
*     12781     52103: contig of 39323 bp in length
*     52104 52203: gap of      100 bp
*     52204     103513: contig of 51310 bp in length
*    103514 103613: gap of      100 bp
*    103614     183215: contig of 79602 bp in length.

```

```

FEATURES             Location/Qualifiers
     source            1. .183215
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-20L19"
                        /clone_lib="RPCI-11 Human Male BAC"
     misc_feature      1. .2147
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:left"
     misc_feature      2248. .3360
                        /note="assembly_fragment"
     misc_feature      3461. .4570
                        /note="assembly_fragment"
     misc_feature      4671. .5920
                        /note="assembly_fragment"
     misc_feature      6021. .7417
                        /note="assembly_fragment"
     misc_feature      7518. .8558
                        /note="assembly_fragment"
     misc_feature      8659. .9869
                        /note="assembly_fragment"
     misc_feature      9970. .12680
                        /note="assembly_fragment"
     misc_feature      12781. .52103
                        /note="assembly_fragment"
     misc_feature      52204. .103513
                        /note="assembly_fragment"
     misc_feature      103614. .183215
                        /note="assembly_fragment"
                        clone_end:T7
                        vector_side:right"

```

```

BASE COUNT      59016 a   32076 c   31329 g   59794 t   1000 others
ORIGIN

```

```

Query Match          8.7%;   Score 36.6;   DB 2;   Length 183215;
Best Local Similarity 60.6%;   Pred. No. 3.3;

```

Matches 60; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy 310 tctctgttcaatcatgctgaaaaacagtgggtgggatttctgggccaatacatatatcg 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19667 TCTATGTTTCTAGATGGGGACCAACTGTGACTAGTTATTTTGGAGCAAATCCAGCTAACT 19608

Qy 370 caggggatagatgaccgagtgatctcgccctcagtggca 408
    || || ||| || || ||||| || ||||| |||
Db 19607 CAAGGAATACTCAAACAGTGATAGGGTCCTCAGTAGCA 19569
```

RESULT 10
AE003825/c
LOCUS AE003825 278196 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386047 section 18
of 52, complete sequence.
ACCESSION AE003825 AE002787
VERSION AE003825.2 GI:10727634
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 278196)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,

Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

TITLE The genome sequence of *Drosophila melanogaster*
JOURNAL Science 287 (5461), 2185-2195 (2000)
MEDLINE 20196006
REFERENCE 2 (bases 1 to 278196)
AUTHORS Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT On Oct 9, 2000 this sequence version replaced gi:7303570.

FEATURES

source	1. .278196 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2R"
mRNA	join(6421. .7995, 36785. .36945, 37365. .37899) /gene="inv" /note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript." /product="CT39612" /db_xref="FLYBASE:FBan0017835" /db_xref="FLYBASE:FBgn0001269"
gene	<6421. .>37899 /gene="inv" /note="CG17835" /map="47F15-48A1" /db_xref="FLYBASE:FBan0017835" /db_xref="FLYBASE:FBgn0001269"
CDS	join(6692. .7995, 36785. .36945, 37365. .37606) /gene="inv" /note="inv gene product; Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript" /codon_start=1 /db_xref="FLYBASE:FBan0017835" /db_xref="FLYBASE:FBgn0001269" /protein_id="AAF58640.1" /db_xref="GI:7303587" /translation="MSTLASTRPPPLKLTIPSLEEAEADHAQERRAGGGGQEVGKMHPD CLPLPLVQPGNSPQVREEEDEQTECEEQLNIEDEEVEEEHDLDDLEDPASCCSENSVL SVGQEQSEAAQAALSAQAQARQLLISQIYRPSAFSSTATTVLPPSEGPFPSPEDLLQ LPPSTGTTFQEEFLRKSQLYAEELMKQMQMHLMAAARVNALTAAGKQLQMAMAAA TVPSGQDALAQLTATALGLPGGAVHPHQQLLLQDQVHHHHMQNHLNNNNLHERA LKFSIDNILKADFGSRLPKIGALSGNIGGSGVSGSSTGSSKNSGTNGNRSPLKAPKK SGKPLNLAQSNAAANSSLSFSSSLANICSNSNSNSTATSSSTNTSGAPVDLVKSP PAAGAGATGASGKSGEDSGTPIVWPAWVYCTRYSDRPSSGESKSTSAKAQEAGDVQFG GRWWGWRREGGRRWGRGAGGQKAANGLQRNAVQTEARVQRESLSDGEATPAAERG TGTERGADQDLVPEQTGQAEKVERHQESAGAAADGAGIVQPLDDTADPRGGAAGAAG GG"
mRNA	complement(join(54356. .54366, 54769. .55563, 55844. .55941, 57075. .58578)) /gene="en"

```

/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT25904"
/db_xref="FLYBASE:FBan0009015"
/db_xref="FLYBASE:FBgn0000577"
gene complement(<54356. .>58578)
/genes="en"
/note="CG9015"
/map="48A1-48A3"
/db_xref="FLYBASE:FBan0009015"
/db_xref="FLYBASE:FBgn0000577"
CDS complement(join(55315. .55563,55844. .55941,57075. .58386))
/genes="en"
/note="en gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0009015"
/db_xref="FLYBASE:FBgn0000577"
/protein_id="AAF58639.1"
/db_xref="GI:7303586"
/translation="MALEDRCSPQSAPSPITLQMQLHLLHQQQQQQQQQQQQMQLHLHQLQ
QLQQLHQQQLAAGVFHHPAMAFDAAAAAAAAAAAAAAAAHAHAALQORLSGSGSPASCS
TPASSTPLTIKEEESDSVIGDMSFHNQTHTTNEEEAEEDDDIDVDVDDTSAGGRLLP
PAHQQQSTAKPSLAFSISNILSDRFGDVQKPGKSMENQASIFRPFEASRSQTATPSAF
TRVDLLEFSRQQQAAAAAATAAMMLERANFLNCFNPAAYPRIHEEIVQSRLRRAANA
VIPPMMSSKMSDANPEKSALGSLCKAVSQIGQPAAPTMTQPPLSSSASSLASPPASN
ASTISSTSSVATSSSSSSSSSGCSSAASSLNSSPSSRLGASGSGVNASSPQPQPIPPPSA
VSRDSGMESSDDTRSETGSTTTTEGGKNEMWPAWVYCTRYSDRPSSGPRYRRPKQPKDK
TNDEKRPRTAFSSQELARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAK
IKKSTGSKNPLALQLMAQGLYNHTTVPLTKEEEELEMRMNGQIP"
mRNA complement(join(<110327. .110528,110589. .110699,
110765. .110886,110951. .111578,111665. .112114,
112344. .112861,113731. .113926,113977. .114194,
114258. .116277,116383. .117741,117813. .117854,
118254. .118298,118362. .118427,118487. .118843,
119079. .119351,119414. .120656,120719. .121341,
121403. .121870,122401. .122646,122709. .122835))
/genes="CG10897"
/product="CT30517"
/db_xref="FLYBASE:FBan0010897"
/db_xref="FLYBASE:FBgn0033636"
gene complement(<110327. .>122835)
/genes="CG10897"
/map="48A3-48A5"
/db_xref="FLYBASE:FBan0010897"
/db_xref="FLYBASE:FBgn0033636"
CDS complement(join(121527. .121870,122401. .122646,
122709. .122835))
/genes="CG10897"
/note="CG10897 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0010897"
/db_xref="FLYBASE:FBgn0033636"
/protein_id="AAF58638.1"
/db_xref="GI:7303585"
/translation="MNKNAGDGS DGKNSNKNNSNAGGGGGAGGPHDPTGLLDAASLFAY

```

WGRDPTGAAAAASNPLFNSQFNAAAAAGLGLLPQAGGASANDRYSMAAAAAAGAH
 HHQNTMAVAASQAASLAGLHPAISCPLLQSPASLGSFPLSTSRSGGCLWTSWDGNGR
 RCWWSGCWRWWSRCTRPGRRWKWRKWQRWWWILIGFFRQQVKQVAQGEASRPAATTAT
 AKSGQQSKCGSCGSSSSSSQ"

mRNA complement(join(147646. .148170,148232. .148556,
 148618. .>149725))
 /gene="CG9006"
 /product="CT25884"
 /db_xref="FLYBASE:FBan0009006"
 /db_xref="FLYBASE:FBgn0033637"

gene complement(<147646. .>149725)
 /gene="CG9006"
 /map="48B1-48B1"
 /db_xref="FLYBASE:FBan0009006"
 /db_xref="FLYBASE:FBgn0033637"

CDS complement(join(147646. .148170,148232. .148556,
 148618. .149687))
 /gene="CG9006"
 /note="CG9006 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0009006"
 /db_xref="FLYBASE:FBgn0033637"
 /protein_id="AAF58637.1"
 /db_xref="GI:7303584"
 /translation="MRPNLFSGASRLITYSRNGKLLTRGRSTKATSSSLDSQHQAAT
 TEGGRAESVEESPEQQRKLPTREPLAKNFFIGVVDKELLAYPEVIPRDEMAQLENSLL
 PLKNYFVEPRETEETSPETLRQLGLYGLNVSTDYEGKGYGWSASLMASEPDSTDINVT
 LGLQTHRNVVDLLKEVGTPLOQQRYLQDLATGKLIGTEAIYEISPPEEDYFNNTAELF
 PEYGKWQLNGEKSFVICTPGERQLFLVLAQTQQPNVPGVLGRGTTIFLVDSQQEGVRL
 GEKHATFGCRKAEIRRVHFEGVKLGEDQVVGLPHDGNRYSEQLVRSSRLRGSVLGLSL
 AKKLLNELAQYTVNTTQCGVQLQDLELTRIHSRAMCSVYAMESMLYLTAGLLDEFRA
 QDVTLESAITKYFTLRQVYAIASQNLGVVGPKSLLSGETTELGLRDAQAQLCTQGESLD
 TLGMFIALTGLQHAGQAMNTGVRKSRNPLFNPGHIFGKFLDNNSIDNPKTKMQLSEHV
 HPSLEAAQCIELSVARLQMAVELMFTKHGNVAVVERQSEMQRLEAVGTLIYAMWASVA
 RASRSYICGLPLADHELLTATAICSEGRDRVRTLCETIYGGHFVNNDNNLVRLSKQVA
 KSKGYFAVHPLTFNF"

mRNA complement(join(151561. .151756,151832. .152133,
 152333. .153187,153247. .154618,154680. .154816,
 154877. .156011,157113. .157127,157294. .157318,
 162657. .>163284))
 /gene="CG9005"
 /product="CT25874"

Query Match 8.7%; Score 36.6; DB 3; Length 278196;
 Best Local Similarity 51.5%; Pred. No. 3.4;
 Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 143 tgtgttcaatgcattctgggaaaggaatgttgagagctctgtgcagccaaggagatgcaa 202
 ||| || || | | ||||| ||| ||| | | ||| ||| | | |||

Db 159599 TGTCTTGAAGTCTTCTGCGAAATGAAAATTGAATAAGACATACAGACAAAAATACAAA 159540

Qy 203 ggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttcagcctatctgaca 262
 | | | | ||||| | | ||| | | ||| | | ||| | | |||

Db 159539 TTAATGCTATTTCAGGCAGCTGTTTGCATCGATTCCGAATAAAGTTTTTACCAATTTAACT 159480

Qy 263 tccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
 | || | || ||| |||| | | | ||| ||

Db 159479 TTAAATGTAATAAAAAATAACTCAAATACTAATAAAGCTTAT 159437

RESULT 11

SAU02975

LOCUS SAU02975 3911 bp mRNA VRT 29-OCT-1993

DEFINITION Squalus acanthias proteolipid protein DM gamma mRNA, complete cds.

ACCESSION U02975

VERSION U02975.1 GI:409973

KEYWORDS .

SOURCE spiny dogfish.

ORGANISM Squalus acanthias

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Elasmobranchii; Squala; Squaloidei; Squalidae; Squalus.

REFERENCE 1 (bases 1 to 3911)

AUTHORS Kitagawa,K., Sinoway,M.P., Yang,C., Gould,R.M. and Colman,D.R.

TITLE A proteolipid protein gene family: expression in sharks and rays
and possible evolution from an ancestral gene encoding a
pore-forming polypeptide

JOURNAL Neuron 11 (3), 433-448 (1993)

MEDLINE 94000810

REFERENCE 2 (bases 1 to 3911)

AUTHORS Colman,D.R.

TITLE Direct Submission

JOURNAL Submitted (28-OCT-1993) David R. Colman, Brookdale Center for
Molecular Biology, Mount Sinai School of Medicine, 1 Gustave L.
Levy Place, New York, NY 10029, USA

FEATURES Location/Qualifiers

source

1. .3911

/organism="Squalus acanthias"

/db_xref="taxon:7797"

/clone="DM beta"

/tissue_type="brain"

/dev_stage="adult"

5'UTR

1. .90

CDS

91. .831

/codon_start=1

/product="proteolipid protein DM gamma"

/protein_id="AAC59641.1"

/db_xref="GI:409974"

/translation="MGCFECCIKCLGGVPYASLLATILCFSGVALFCGCGHVALTKVE
RIVQLYFSNNASDHVLLTDVIQMMHYVIYGVASFSLYGIILLAEGFYTTSAVKEIHG
EFKTTVCGRCISGMSVFLTYLLGIAWLGVFGFSAPPAFIYYNMWSACQTISSPPVNL
TVIEEICVDVRQYGIIPWNASPGKACGSTLTITCNTSEFDLSYHLFIVACAGAGATVI
ALLIYMMATTYNFAVLKFKSREDCCTKF"

3'UTR

832. .3911

BASE COUNT 1201 a 777 c 765 g 1168 t

ORIGIN

Query Match 8.6%; Score 36.4; DB 5; Length 3911;

Best Local Similarity 53.5%; Pred. No. 2.6;

Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 154 cattctgggaaaggaatgttgcagagtctgtgcagccaaggagatgcaaggccatttgtg 213

| | | | | | | | | | | | | | | | | | | | | |

Db 345 CCTCCTGGCAGAAGGGTTTTACACCACAAGCGCTGTTAAGGAGATTCATGGTGAGTTCAA 404

Qy 214 gacgaagctgtgctgcaagtatctgactggggtttcagcctatctgacatccaactgcag 273
 || | ||| | | | || | | | | | | | | |
 Db 405 AACAACTGTGTGTGGACGCTGCATCAGTGAATGTCTGTCTTTCTGACCTACCTGTTGGG 464

Qy 274 aagaaagaggctcaaggctttt 295
 || | ||| | |||
 Db 465 AATAGCGTGGCTGGGAGTTTTT 486

RESULT 12

AL589684/c

LOCUS AL589684 94555 bp DNA PRI 06-APR-2001
 DEFINITION Human DNA sequence from clone RP11-437J19 on chromosome 6, complete sequence.
 ACCESSION AL589684
 VERSION AL589684.7 GI:13561020
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM 'Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 94555)
 AUTHORS Dunn,M.
 TITLE Direct Submission
 JOURNAL Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Apr 8, 2001 this sequence version replaced gi:13398844. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-437J19 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-437J19 It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.
The true left end of clone RP1-124C6 is at 94456 in this sequence.
The true right end of clone RP1-132N8 is at 100 in this sequence.

FEATURES	Location/Qualifiers
source	1. .94555 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /clone="RP11-437J19" /clone_lib="RPCI-11.2"
repeat_region	5. .224 /note="MER1B repeat: matches 1. .222 of consensus"
repeat_region	225. .469 /note="L1MC3 repeat: matches 7460. .7722 of consensus"
repeat_region	1509. .2925 /note="L1M4 repeat: matches 3213. .4663 of consensus"
repeat_region	2927. .3676 /note="L1PA7 repeat: matches 5358. .6137 of consensus"
repeat_region	3699. .5060 /note="L1MEc repeat: matches 2212. .3224 of consensus"
repeat_region	5298. .5729 /note="L1M4c repeat: matches 1546. .1973 of consensus"
repeat_region	5812. .6559 /note="L1M4c repeat: matches 615. .1388 of consensus"
repeat_region	6560. .6952 /note="MSTA repeat: matches 1. .426 of consensus"
repeat_region	6953. .7256 /note="L1M4c repeat: matches 318. .615 of consensus"
repeat_region	7334. .7730 /note="L1MB8 repeat: matches 5760. .6165 of consensus"
repeat_region	9532. .9885 /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region	11134. .11181 /note="24 copies 2 mer ta 83% conserved"
repeat_region	11138. .11181 /note="11 copies 4 mer tata 86% conserved"
repeat_region	11182. .11255 /note="37 copies 2 mer at 71% conserved"
repeat_region	11218. .11257 /note="10 copies 4 mer atat 90% conserved"
repeat_region	13859. .16586 /note="L1PA13 repeat: matches 3426. .6153 of consensus"
repeat_region	16647. .16971 /note="L2 repeat: matches 2340. .2750 of consensus"
repeat_region	17554. .17737 /note="MIR repeat: matches 49. .233 of consensus"
repeat_region	18069. .18376 /note="AluSx repeat: matches 1. .308 of consensus"
repeat_region	18474. .18699 /note="MIR repeat: matches 11. .261 of consensus"
repeat_region	18842. .18885 /note="11 copies 4 mer caca 88% conserved"
repeat_region	19896. .20063 /note="MER77 repeat: matches 445. .631 of consensus"
repeat_region	20073. .20108 /note="MER77 repeat: matches 274. .308 of consensus"
repeat_region	20109. .20345

repeat_region	/note="MER46A repeat: matches 1. .235 of consensus"
	20346. .20435
repeat_region	/note="MER77 repeat: matches 181. .274 of consensus"
	21689. .22060
repeat_region	/note="L1PA15 repeat: matches 5779. .6153 of consensus"
	22082. .22117
repeat_region	/note="9 copies 4 mer aaat 86% conserved"
	22282. .23193
repeat_region	/note="L1MC2 repeat: matches 5404. .6325 of consensus"
	23308. .23329
repeat_region	/note="11 copies 2 mer ca 100% conserved"
	23365. .23427
repeat_region	/note="L2 repeat: matches 2639. .2695 of consensus"
	24523. .24800
repeat_region	/note="MIR repeat: matches 4. .262 of consensus"
	24823. .25274
repeat_region	/note="MER74B repeat: matches 176. .622 of consensus"
	25326. .25438
repeat_region	/note="MER74A repeat: matches 1. .113 of consensus"
	26031. .26221
repeat_region	/note="MER5A repeat: matches 2. .189 of consensus"
	27353. .27442
repeat_region	/note="L1PA4 repeat: matches 6055. .6144 of consensus"
	28065. .28092
repeat_region	/note="14 copies 2 mer tg 92% conserved"
	28811. .28859
repeat_region	/note="L2 repeat: matches 2700. .2745 of consensus"
	28866. .29053
repeat_region	/note="MLT1J repeat: matches 1. .191 of consensus"
	29158. .29277
repeat_region	/note="60 copies 2 mer ag 64% conserved"
	29279. .29708
repeat_region	/note="MLT2CB repeat: matches 1. .453 of consensus"
	29715. .29840
repeat_region	/note="MLT1J repeat: matches 385. .516 of consensus"
	30941. .31688
repeat_region	/note="L1M4 repeat: matches 4877. .5642 of consensus"
	32014. .32167
repeat_region	/note="77 copies 2 mer tt 74% conserved"
	32019. .32154
repeat_region	/note="4 copies 34 mer 76% conserved"
	32045. .32108
repeat_region	/note="16 copies 4 mer ttct 100% conserved"
	32119. .32170
repeat_region	/note="13 copies 4 mer ttct 98% conserved"
	32170. .32477
repeat_region	/note="AluSc repeat: matches 1. .309 of consensus"
	32915. .32954
repeat_region	/note="20 copies 2 mer ac 82% conserved"
	33280. .33583
repeat_region	/note="AluJb repeat: matches 3. .303 of consensus"
	34082. .34285
repeat_region	/note="6 copies 34 mer 71% conserved"
	34083. .34274
repeat_region	/note="48 copies 4 mer aaag 71% conserved"
	35357. .35409
repeat_region	/note="MER5A repeat: matches 16. .65 of consensus"

repeat_region 35499. .35516
 /note="MER5A repeat: matches 65. .184 of consensus"
 repeat_region 35547. .36313
 /note="L1MEc repeat: matches 583. .1464 of consensus"
 repeat_region 38689. .38890
 /note="L2 repeat: matches 2544. .2750 of consensus"
 repeat_region 39096. .39409
 /note="AluY repeat: matches 1. .311 of consensus"
 repeat_region 39875. .40054
 /note="MLT1D repeat: matches 52. .218 of consensus"
 repeat_region 40055. .40409
 /note="THE1B repeat: matches 1. .364 of consensus"
 repeat_region 40410. .40651
 /note="MLT1D repeat: matches 218. .503 of consensus"
 repeat_region 40936. .41082
 /note="L2 repeat: matches 2560. .2708 of consensus"
 repeat_region 42602. .42792
 /note="MER5A repeat: matches 7. .187 of consensus"
 repeat_region 42845. .42940
 /note="L2 repeat: matches 2611. .2750 of consensus"
 repeat_region 42955. .43071
 /note="MER45 repeat: matches 10. .120 of consensus"
 repeat_region 43357. .43527
 /note="MER5A repeat: matches 1. .189 of consensus"
 repeat_region 44098. .44231
 /note="67 copies 2 mer aa 58% conserved"
 repeat_region 45604. .45714
 /note="L1PA13 repeat: matches 6047. .6156 of consensus"
 repeat_region 45760. .45911
 /note="76 copies 2 mer aa 57% conserved"
 repeat_region 47006. .47308
 /note="AluSx repeat: matches 1. .302 of consensus"
 repeat_region 47889. .47982
 /note="MER81 repeat: matches 20. .113 of consensus"
 repeat_region 48733. .48764
 /note="MLT1J repeat: matches 108. .139 of consensus"
 repeat_region 49237. .49306
 /note="MER31-internal repeat: matches 4819. .4886 of consensus"
 repeat_region 49311. .49747
 /note="MER31B repeat: matches 1. .450 of consensus"

Query Match 8.6%; Score 36.2; DB 9; Length 94555;
 Best Local Similarity 50.9%; Pred. No. 4.1;
 Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 216 cgaagctgtgctgcaagtatctgactgggggtttcagcctatctgacatccaactgcagaa 275
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21249 CAAATCAGAGAAGCTAGTGTGTCAGAAGAGGTTTTAGGAAGGTGAAGATTAGATGTGCAGGC 21190

 Qy 276 gaaagaggctcaaggcttttttgaactcatcacgtctctgttcaatcatgctgaaaaaca 335
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21189 GGATGAGGCTAGAGCCAAATAAGCAGATTACTAGGTTATGTGTGCTTATGGGGAAAGGCA 21130

 Qy 336 gtggggtgggattttctggggcccaatacatatatcgaggggatagatgac 384
 | | | | | | | | | | | | | | | | | | | | | |
 Db 21129 GTGGCAGGAAAATTGGGTGCCAAGAGATAAATGCCAGCCATTGGCAAC 21081

RESULT 13
 AC024632/c
 LOCUS AC024632 168438 bp DNA HTG 03-MAR-2000
 DEFINITION Homo sapiens chromosome 6 clone RP11-437J19 map 6, WORKING DRAFT
 SEQUENCE, 25 unordered pieces.
 ACCESSION AC024632
 VERSION AC024632.1 GI:7139757
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 168438)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 6, clone RP11-437J19
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 168438)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6062
 Center clone name: 437_J_19
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 154266 bases at least Q40
Consensus quality: 161752 bases at least Q30
Consensus quality: 164265 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 166038; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1505: contig of 1505 bp in length
* 1506 1605: gap of 100 bp
* 1606 3527: contig of 1922 bp in length
* 3528 3627: gap of 100 bp
* 3628 5122: contig of 1495 bp in length
* 5123 5222: gap of 100 bp
* 5223 8133: contig of 2911 bp in length
* 8134 8233: gap of 100 bp
* 8234 9885: contig of 1652 bp in length
* 9886 9985: gap of 100 bp
* 9986 12052: contig of 2067 bp in length
* 12053 12152: gap of 100 bp
* 12153 14403: contig of 2251 bp in length
* 14404 14503: gap of 100 bp
* 14504 18869: contig of 4366 bp in length
* 18870 18969: gap of 100 bp
* 18970 21983: contig of 3014 bp in length
* 21984 22083: gap of 100 bp
* 22084 26566: contig of 4483 bp in length
* 26567 26666: gap of 100 bp
* 26667 30100: contig of 3434 bp in length
* 30101 30200: gap of 100 bp
* 30201 33413: contig of 3213 bp in length
* 33414 33513: gap of 100 bp
* 33514 36961: contig of 3448 bp in length
* 36962 37061: gap of 100 bp
* 37062 42449: contig of 5388 bp in length
* 42450 42549: gap of 100 bp
* 42550 47375: contig of 4826 bp in length
* 47376 47475: gap of 100 bp
* 47476 51908: contig of 4433 bp in length
* 51909 52008: gap of 100 bp
* 52009 57797: contig of 5789 bp in length
* 57798 57897: gap of 100 bp
* 57898 64548: contig of 6651 bp in length
* 64549 64648: gap of 100 bp
* 64649 72958: contig of 8310 bp in length
* 72959 73058: gap of 100 bp
* 73059 85343: contig of 12285 bp in length
* 85344 85443: gap of 100 bp


```

misc_feature      73059. .85343
                  /note="assembly_fragment"
misc_feature      85444. .97769
                  /note="assembly_fragment"
misc_feature      97870. .108295
                  /note="assembly_fragment"
misc_feature      108396. .123953
                  /note="assembly_fragment"
misc_feature      124054. .142803
                  /note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature      142904. .168438
                  /note="assembly_fragment"
BASE COUNT      52247 a 30934 c 30484 g 52372 t 2401 others
ORIGIN

```

```

Query Match      8.6%; Score 36.2; DB 2; Length 168438;
Best Local Similarity 50.9%; Pred. No. 4.3;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

Qy 216 cgaagctgtgctgcaagtatctgactgggggtttcagcctatctgacatccaactgcagaa 275
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161456 CAAATCAGAGAAGCTAGTGTCAGAAGAGGTTTTAGGAAGGTGAAGATTAGATGTGCAGGC 161397

Qy 276 gaaagaggctcaaggcttttttgaactcatcacgtctctgttcaatcatgctgaaaaaca 335
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161396 GGATGAGGCTAGAGCCAAATAAGCAGATTACTAGGTTATGTGTGCTTATGGGGAAAGGCA 161337

Qy 336 gtgggtgggatttctgggcccatacatatatcgaggggatagatgac 384
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161336 GTGGCAGGAAAATTGGGTGCCAAGAGATAAATGCCCAGCCATTGGCAAC 161288

```

```

RESULT 14
AF137266
LOCUS      AF137266      824 bp      DNA      PLN      09-JAN-2001
DEFINITION Nuphar lutea CT dinucleotide repeat microsatellite sequence.
ACCESSION  AF137266
VERSION    AF137266.1 GI:5733431
KEYWORDS   .
SOURCE     Nuphar lutea.
ORGANISM   Nuphar lutea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Nymphaeaceae; Nuphar.
REFERENCE  1 (bases 1 to 824)
AUTHORS    Ouborg,N.J., Goodall-Copestake,W.P., Saumitou-Laprade,P., Bonnin,I.
            and Epplen,J.T.
TITLE      Novel polymorphic microsatellite loci isolated from the yellow
            waterlily, Nuphar lutea
JOURNAL    Mol. Ecol. 9 (4), 497-498 (2000)
MEDLINE    20200292
PUBMED     10736057
REFERENCE  2 (bases 1 to 824)
AUTHORS    Ouborg,N.J., Goodall-Copestake,W.P., Saumitou-Laprade,P. and
            Epplen,J.T.

```

TITLE Direct Submission
 JOURNAL Submitted (25-MAR-1999) Dept. of Ecology, Univ. of Nijmegen,
 Toernooiveld 1, Nijmegen 6525 ED, Netherlands
 FEATURES Location/Qualifiers
 source 1. .824
 /organism="Nuphar lutea"
 /db_xref="taxon:77113"
 repeat_region 271. .316
 /note="microsatellite"
 /rpt_type=tandem
 /rpt_unit=CT
 BASE COUNT 223 a 193 c 163 g 228 t 17 others
 ORIGIN

Query Match 8.6%; Score 36; DB 8; Length 824;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 54; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 260 acatccaactgcagaagaaagaggctcaaggcttttttgaactcatcacgtctctgttca 319
 ||||||| | ||| ||||| || | || || |||| | | |
 Db 345 ACATCCAACGGTAGATATCAGAGGCTCAAGCCTGATGAAAAATCGTCACGGCCAAGAAGA 404

 Qy 320 atcatgctgaaaaacagtgggtgg 343
 | |||| || || || ||
 Db 405 AGATGGCTGAGAATCAAAGGGAGG 428

RESULT 15
 AC010628
 LOCUS AC010628 101261 bp DNA HTG 18-JUL-2000
 DEFINITION Homo sapiens chromosome 5 clone CTD-2180L11, WORKING DRAFT
 SEQUENCE, 16 ordered pieces.
 ACCESSION AC010628
 VERSION AC010628.4 GI:9256262
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 101261)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 5
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 101261)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Jul 18, 2000 this sequence version replaced gi:7710608.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

 Project Information
 Center Project Name: 696944, H468

Center clone name: CITB-H1_2180L11

Summary Statistics

Consensus quality: 76583 bases at least Q40

Consensus quality: 90678 bases at least Q30

Consensus quality: 96052 bases at least Q20

Estimated insert size: 112000; pulse field gel estimation

Estimated insert size: 100561; sum-of-contigs estimation

Quality coverage: 3.95 in Q20 bases; pulse field gel estimation

Quality coverage: 4.4 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 16 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 2349: contig of 2349 bp in length

* 2350 2449: gap of unknown length

* 2450 9244: contig of 6795 bp in length

* 9245 9344: gap of unknown length

* 9345 12209: contig of 2865 bp in length

* 12210 12309: gap of unknown length

* 12310 14574: contig of 2265 bp in length

* 14575 14674: gap of unknown length

* 14675 16514: contig of 1840 bp in length

* 16515 16614: gap of unknown length

* 16615 17989: contig of 1375 bp in length

* 17990 18089: gap of unknown length

* 18090 37668: contig of 19579 bp in length

* 37669 37768: gap of unknown length

* 37769 41885: contig of 4117 bp in length

* 41886 41985: gap of unknown length

* 41986 51375: contig of 9390 bp in length

* 51376 51475: gap of unknown length

* 51476 54268: contig of 2793 bp in length

* 54269 54368: gap of unknown length

* 54369 55918: contig of 1550 bp in length

* 55919 56018: gap of unknown length

* 56019 58548: contig of 2530 bp in length

* 58549 58648: gap of unknown length

* 58649 64654: contig of 6006 bp in length

* 64655 64754: gap of unknown length

* 64755 76433: contig of 11679 bp in length

* 76434 76533: gap of unknown length

* 76534 82850: contig of 6317 bp in length

* 82851 82950: gap of unknown length

* 82951 101261: contig of 18311 bp in length.

FEATURES

source

Location/Qualifiers

1. .101261

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/clone="CTD-2180L11"

/clone_lib="CalTech human BAC library D"

BASE COUNT	29116 a	17494 c	19600 g	33526 t	1525 others
ORIGIN					

Query Match 8.5%; Score 35.6; DB 2; Length 101261;
Best Local Similarity 54.6%; Pred. No. 6.3;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 28 ttagcttaacaattccttagtagtcaccccttcgattaaatgtcaacatttgccttttcgc 87
 || || || || ||| || | || |||| | | || | |||
Db 40727 TTTGCATAGAAAAATTATTATTTGTATATCTCATAAATACCTAAATCTCATTTTCAA 40786

Qy 88 gttccaattactaatgttacggcattattcaggacagaactttactggaacgtcctgtgt 147
 ||| | || |||| |||| || || |||| | | |||| || |||| |
Db 40787 ATTCTATTTGTTAATTTTACTAAATGATGATATACAGTATTGTACTTTAAGACCCTGATT 40846

Qy 148 tcaatgcatt 157
 | ||| |||
Db 40847 CCTATGAATT 40856

Search completed: February 7, 2002, 10:55:05
Job time: 9231 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      February  7, 2002, 10:59:33 ; Search time 428.31 Seconds
              (without alignments)
              842.693 Million cell updates/sec
```

```
Title:          US-09-394-745-5950
Perfect score:  421
Sequence:       1 ggggtccaggcacgcgtccga.....agtggcagaattttgtgccgc 421
```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :      N_Geneseq_1101:*
1:  /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2:  /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3:  /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
```

4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
 5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
 6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
 7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
 8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
 9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
 10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
 11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
 12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
 13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
 14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
 15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
 16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
 17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
 18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
 19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
 20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
 21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
 22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length	Match	Length			
	1	34.8	8.3	936	22	AAF58252			Oligonucleotide D1
	2	34.8	8.3	936	22	AAF58254			Oligonucleotide D1
	3	34.8	8.3	936	22	AAF58257			Oligonucleotide D1
	4	34.8	8.3	936	22	AAF58259			Oligonucleotide D2
	5	34.8	8.3	936	22	AAF58262			Oligonucleotide D2
	6	34.8	8.3	938	22	AAF58255			Oligonucleotide D1
c	7	34.2	8.1	786	22	AAH07451			Human cDNA clone (
c	8	34.2	8.1	1753	22	AAH16513			Human cDNA sequenc
c	9	33.6	8.0	936	22	AAF58252			Oligonucleotide D1
c	10	33.6	8.0	936	22	AAF58254			Oligonucleotide D1
c	11	33.6	8.0	936	22	AAF58257			Oligonucleotide D1
c	12	33.6	8.0	936	22	AAF58259			Oligonucleotide D2
c	13	33.6	8.0	936	22	AAF58262			Oligonucleotide D2
c	14	33.6	8.0	938	22	AAF58255			Oligonucleotide D1
	15	33	7.8	752	20	AAX98756			Human validated ca
c	16	32.6	7.7	878	22	AAH07610			Human cDNA clone (
c	17	32.6	7.7	2187	22	AAH14871			Human cDNA sequenc
c	18	32.6	7.7	7418	22	AAI58480			Human polynucleoti
c	19	31	7.4	2878	15	AAQ54482			Excitatory amino a
c	20	31	7.4	2878	16	AAQ91232			Human EAA4 recepto
c	21	31	7.4	2878	22	AAC62038			cDNA encoding a un
c	22	31	7.4	2878	22	AAC62041			cDNA encoding a fo
c	23	31	7.4	2878	22	AAC62042			cDNA encoding a fo
c	24	31	7.4	236303	22	AAS11614			Human genomic DNA
	25	30.6	7.3	547	22	AAF68173			Human lung tumour
c	26	30.6	7.3	1110	21	AAC45497			Arabidopsis thalia
	27	30.2	7.2	700	22	AAH92694			Human inflammatory
	28	30.2	7.2	3975	21	AAC51553			Arabidopsis thalia

	29	30	7.1	1366	21	AAA01986	Human colon cancer
	30	30	7.1	1575	18	AAT91305	Human J59-41 secre
c	31	30	7.1	2400	14	AAQ42496	Full-length human
c	32	30	7.1	2416	21	AAZ49470	Human wild type Bu
c	33	30	7.1	2416	21	AAZ49471	Human Butyrylcholi
c	34	29.8	7.1	391	20	AAZ28002	Nucleotide sequenc
c	35	29.8	7.1	391	21	AAA96029	Human lectomedin-1
c	36	29.6	7.0	1038602	20	AAZ01425	Complete genome se
c	37	29.4	7.0	963	22	AAH31657	Human olfactory re
c	38	29.4	7.0	3935	21	AAZ89458	Murine trans-synap
c	39	29.4	7.0	1664976	19	AAV21209	Methanococcus jann
	40	29.2	6.9	1013	21	AAZ51231	Staphylococcus aur
	41	29.2	6.9	8779	18	AAV74369	Staphylococcus aur
c	42	29.2	6.9	89047	22	AAF28547	Genomic fragment #
c	43	29	6.9	1112	21	AAC40621	Arabidopsis thalia
	44	29	6.9	1544	22	AAH02939	Human shear stress
	45	29	6.9	2204	16	AAQ87426	Human GRK cDNA #2.

ALIGNMENTS

RESULT 1

AAF58252

ID AAF58252 standard; DNA; 936 BP.

XX

AC AAF58252;

XX

DT 24-APR-2001 (first entry)

XX

DE Oligonucleotide D1835.

XX

KW Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.

XX

OS Synthetic.

XX

PN WO200107665-A2.

XX

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-US20476.

XX

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX

PI Umek RM;

XX

DR WPI; 2001-159728/16.

XX

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -

XX

PS Example 6; Page 127; 159pp; English.

PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 8.3%; Score 34.8; DB 22; Length 936;
Best Local Similarity 1.7%; Pred. No. 0.18;
Matches 6; Conservative 195; Mismatches 147; Indels 0; Gaps 0;

[illegible]

[illegible]

RESULT 4

AAAF58259

ID AAF58259 standard; DNA; 936 BP.

XX

AC AAF58259;

XX

DT 24-APR-2001 (first entry)

XX

DE Oligonucleotide D2004.

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX

OS Synthetic.

XX

PN WO200107665-A2.

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

XX

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -

XX

PS Example 6; Page 128; 159pp; English.

XX

CC acids each containing an electron-transfer group (ETM) having

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX

SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 8.3%; Score 34.8; DB 22; Length 936;
Best Local Similarity 1.7%; Pred. No. 0.18;
Matches 6; Conservative 195; Mismatches 147; Indels 0; Gaps 0;

```
Qy 20 aattgaggtagcttaacaattcttagtagtcaccccttcgattaaatgtcaacatttgc 79
    :|:: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 wwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 251

Qy 80 ctttgcgcttccaattactaatgttacggcattattcaggacagaactttactggaacg 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 311

Qy 140 tcctgtgttcaatgcattctgggaaaggaatgttgagagtcgtgcagccaaggagatg 199
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 371

Qy 200 caaggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttcagcctatctg 259
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 431

Qy 260 acatccaactgcagaagaaagaggtcaaggcttttttgaactcatcacgtctctgttca 319
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 491

Qy 320 atcatgctgaaaaacagtgggtgggatttctgggcccaatacatatat 367
    :|:: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 539
```

RESULT 5

AAF58262

ID AAF58262 standard; DNA; 936 BP.

XX

AC AAF58262;

XX

DT 24-APR-2001 (first entry)

XX

DE Oligonucleotide D2007.

XX

KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.

XX

OS Synthetic.

XX

PN WO200107665-A2.

XX

PD 01-FEB-2001.

XX

PF 26-JUL-2000; 2000WO-US20476.


```

      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    312 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 371
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy    200 caaggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttcagcctatctg 259
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    372 wwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwww 431
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy    260 acatccaactgcagaagaaagaggctcaaggcttttttgaactcatcacgtctctgttca 319
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    432 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 491
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy    320 atcatgctgaaaaacagtgggtgggatttctgggcccaatacatatat 367
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db    492 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 539

```

```

RESULT 7
AAH07451/c
ID AAH07451 standard; cDNA; 786 BP.
XX
AC AAH07451;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4286.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ.ID 4286; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602

```

full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 786 BP; 252 A; 156 C; 161 G; 211 T; 6 other;

Query Match 8.1%; Score 34.2; DB 22; Length 786;
Best Local Similarity 58.3%; Pred. No. 0.26;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 195 agatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttcagcct 254

Qy 255 atctgacatccaactgcagaagaaagaggctcaaggctttttt 297

Db 235 ATCTCCCATTAACTTTTTAAGTAGGGAGCCAACTGTTTCTTT 193

RESULT 8

ID AAH16513 standard; cDNA; 1753 BP.

AC AAH16513;

DT

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:15552.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.


```

RESULT 12
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

```

[illegible]

XX
 AC AAX98756;
 XX
 DT 24-SEP-1999 (first entry)
 XX
 DE Human validated cancer cell derived cDNA #78.
 XX
 KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
 KW integral membrane protein; aspartyl protease; GATA family; wnt family;
 KW transcription factor; G-protein alpha subunit; protein phosphatase;
 KW phorbol ester binding protein; diacylglycerol binding protein; trypsin;
 KW protein kinase; tyrosine phosphatase; developmental signalling protein;
 KW WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
 KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
 KW Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
 KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
 KW prostate; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9933982-A2.
 XX
 PD 08-JUL-1999.
 XX
 PF 22-DEC-1998; 98WO-US27610.
 XX
 PR 21-DEC-1998; 98US-0217471.
 PR 23-DEC-1997; 97US-0068755.
 PR 03-APR-1998; 98US-0080664.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-430243/36.
 XX
 PT New isolated human polynucleotides
 XX
 PS Claim 1; Page 444-445; 591pp; English.
 XX
 CC This invention describes novel isolated human polynucleotides obtained
 CC by screening for differential expression in colon cancer, breast cancer
 CC and lung cancer cell lines. The polynucleotides of the invention are
 CC represented in AAX98275-X99118 and encode polypeptides of protein
 CC families selected from 4 transmembrane segments integral membrane
 CC proteins, 7 transmembrane receptors, ATPases associated with various
 CC cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
 CC transcription factors, G-protein alpha subunit, phorbol esters or
 CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
 CC protein tyrosine phosphatase, trypsin, wnt family of developmental

CC signalling proteins and WW/rsp5/WWP domain containing proteins. The
 CC encoded polypeptides also have a functional domain selected from Ank
 CC repeat, basic region plus leucine zipper transcription factors,
 CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
 CC (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease
 CC domain. The polynucleotides encode polypeptides with similarity to known
 CC protein families and are predicted to have similar properties. The novel
 CC polynucleotides can be used to develop products for use as therapeutic
 CC agents and in forensics, genetic analysis, mapping and diagnostic
 CC applications. In particular, the product can be used for the detection
 CC and management of cancers. They can be used for treating e.g. cervical
 CC cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
 CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
 CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
 CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
 CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,
 CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
 CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
 CC prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
 CC the skin.

XX

SQ Sequence 752 BP; 204 A; 191 C; 166 G; 173 T; 18 other;

Query Match 7.8%; Score 33; DB 20; Length 752;
 Best Local Similarity 67.2%; Pred. No. 0.63;
 Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 141 cctgtgttcaatgcattctgggaaggaatgttgagagtcgtgcagccaaggagatgc 200
 ||||| || ||| | ||||| || || ||| || | ||||| ||||| | |
 Db 540 cctgtaatcccagcactttgggaagcaaangtggcaggatcattccagcccaggagatttc 599
 Qy 201 aaggcca 207
 ||| ||
 Db 600 aaganca 606

Search completed: February 7, 2002, 10:59:37
 Job time: 4963 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 10:51:34 ; Search time 172.96 Seconds
 (without alignments)
 551.268 Million cell updates/sec

Title: US-09-394-745-5950
 Perfect score: 421
 Sequence: 1 ggggtccaggcacgcgtccga.....agtggcagaattttgtgccgc 421

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			%	Query			ID	Description
	No.	Score	Match	Length	DB			
c	1	31	7.4	2878	1	US-07-903-456-1		Sequence 1, Appli
c	2	31	7.4	2878	3	US-08-666-221B-5		Sequence 5, Appli
c	3	31	7.4	2878	3	US-08-666-221B-11		Sequence 11, Appl
c	4	31	7.4	2878	3	US-08-666-221B-13		Sequence 13, Appl
	5	30.6	7.3	2646	1	US-08-539-304A-5		Sequence 5, Appli
c	6	30.6	7.3	6253	2	US-08-627-151A-5		Sequence 5, Appli
c	7	30	7.1	2381	2	US-08-318-826A-9		Sequence 9, Appli
c	8	30	7.1	2400	6	5215909-13		Patent No. 5215909
c	9	30	7.1	2416	2	US-08-318-826A-8		Sequence 8, Appli
c	10	30	7.1	2416	4	US-09-334-489-1		Sequence 1, Appli
c	11	30	7.1	2416	4	US-09-334-489-2		Sequence 2, Appli
	12	29	6.9	2204	1	US-08-221-817-12		Sequence 12, Appl
	13	29	6.9	2204	1	US-08-454-439-12		Sequence 12, Appl
	14	29	6.9	2204	5	PCT-US94-10487-12		Sequence 12, Appl
	15	29	6.9	2206	1	US-08-221-817-10		Sequence 10, Appl
	16	29	6.9	2206	1	US-08-454-439-10		Sequence 10, Appl
	17	29	6.9	2206	5	PCT-US94-10487-10		Sequence 10, Appl
	18	29	6.9	2848	4	US-08-464-954A-2		Sequence 2, Appli
	19	28.8	6.8	2447	2	US-09-014-969-14		Sequence 14, Appl
c	20	28.8	6.8	9636	1	US-08-323-170B-1		Sequence 1, Appli
c	21	28.6	6.8	221	1	US-07-792-525B-1		Sequence 1, Appli
c	22	28.6	6.8	1915	3	US-09-120-365-2		Sequence 2, Appli
c	23	28.6	6.8	1915	4	US-09-515-039-2		Sequence 2, Appli
c	24	28.6	6.8	1916	3	US-09-120-365-88		Sequence 88, Appl
c	25	28.6	6.8	1916	4	US-09-515-039-88		Sequence 88, Appl
c	26	27.8	6.6	213	5	PCT-US95-13333-1		Sequence 1, Appli
c	27	27.8	6.6	1776	1	US-08-464-523B-4		Sequence 4, Appli
c	28	27.6	6.6	1207	2	US-08-694-869-3		Sequence 3, Appli

c	29	27.6	6.6	1207	3	US-09-349-546-3	Sequence 3, Appli
c	30	27.6	6.6	1301	2	US-08-694-869-4	Sequence 4, Appli
c	31	27.6	6.6	1301	3	US-09-349-546-4	Sequence 4, Appli
c	32	27.6	6.6	1422	2	US-08-694-869-5	Sequence 5, Appli
c	33	27.6	6.6	1422	3	US-09-349-546-5	Sequence 5, Appli
c	34	27.6	6.6	7568	2	US-08-694-869-2	Sequence 2, Appli
c	35	27.6	6.6	7568	3	US-09-349-546-2	Sequence 2, Appli
c	36	27.6	6.6	14636	4	US-09-173-914-6	Sequence 6, Appli
c	37	27.4	6.5	775	4	US-08-998-416-669	Sequence 669, App
	38	27.4	6.5	1776	1	US-08-318-947A-3	Sequence 3, Appli
	39	27.4	6.5	1776	2	US-08-795-303-3	Sequence 3, Appli
	40	27.2	6.5	1772	2	US-08-960-022-13	Sequence 13, Appl
c	41	27.2	6.5	2875	3	US-08-714-918-8	Sequence 8, Appli
c	42	27.2	6.5	2875	4	US-09-265-315-8	Sequence 8, Appli
c	43	27.2	6.5	2875	4	US-09-265-315-8	Sequence 8, Appli
c	44	27.2	6.5	2875	4	US-09-266-417-8	Sequence 8, Appli
c	45	27	6.4	3336	2	US-08-977-554-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
 US-07-903-456-1/c
 ; Sequence 1, Application US/07903456
 ; Patent No. 5574144
 ; GENERAL INFORMATION:
 ; APPLICANT: KAMBOJ, Rajender
 ; APPLICANT: ELLIOTT, Candace
 ; APPLICANT: NUTT, Stephen
 ; TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF
 ; TITLE OF INVENTION: THE EAA4 FAMILY
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 1800 Diagonal Road, Suite 500
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22313-0299
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/903,456
 ; FILING DATE: 19920624
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16777/183/ALLE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703)836-9300
 ; TELEFAX: (703)683-4109
 ; TELEX: 899149

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2878 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 134..226
; FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: 227..2860
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 134..2860
US-07-903-456-1

```

```

Query Match          7.4%; Score 31; DB 1; Length 2878;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

```

Qy      60 gattaaatgtcaacatttgccttttcgcgttccaattactaatgttacggcattattcag 119
        ||||| ||| | ||||| || | || | || ||| | | |||| ||
Db      814 GATTACATGAAACTCCTTGCCTCTTTTCATTTCTTTTAGTAAGGGTTTTGCATCCTTTGT 755

Qy     120 gacagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgagag 179
        ||| | | |||| | || || | | | || ||| ||| |||
Db      754 ATCAGCAGGTAAGTACGAATTTTGAGTCGAAGATTATACCTTGATGGAGCTTTGATGAG 695

Qy     180 tctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctga 239
        |||| | | || || || || | || || || || | |
Db      694 CTCTTGCAAACGAATGAGACCAGTGCTGTCATCATACACAACCGTGACGGTTTTCCACTT 635

Qy     240 ctgggggtttcagcctatctgacatccaactgcagaagaaagag 282
        | || | || | || || || || || || || || || ||
Db      634 GAAAAACTGCACCAGGTCTAAAATGGCACGGCTGAGTGAAGAG 592

```

```

RESULT 2
US-08-666-221B-5/c
; Sequence 5, Application US/08666221B
; Patent No. 6136544
; GENERAL INFORMATION:
; APPLICANT: Kamboj, Rajender
; APPLICANT: Nutt, Stephen
; TITLE OF INVENTION: GLUTAMATE RECEPTOR (OR EAA RECEPTOR)
; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/666,221B
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016777/0308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2878 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 134..226
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 227..2860
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 134..2860
US-08-666-221B-5

```

```

Query Match          7.4%; Score 31; DB 3; Length 2878;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

```

Qy      60 gattaaatgtcaacatttgccttttcgcgttccaattactaatgttacggcattattcag 119
      ||||| ||| | ||||| || | || | || ||| | | |||| ||
Db      814 GATTACATGAAACTCCTTGCCTCTTTTCATTTCTTTTAGTAAGGGTTTTGCATCCTTTGT 755

Qy     120 gacagaactttactggaacgtcctgtgttcaatgcattctgaggaaaggaatgttgagag 179
      ||| | | |||| | || || | | | || | ||| |||
Db      754 ATCAGCAGGTAACCTGACGAATTTTGAGTCGAAGATTATACCTTGATGGAGCTTTGATGAG 695

Qy     180 tctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctga 239
      |||| | | || || | | || | | ||| | |
Db      694 CTCTTGCAAACGAATGAGACCAGTGCTGTTCATCATACACAACCGTGACGGTTTTCCACTT 635

Qy     240 ctgggggtttcagcctatctgacatccaactgcagaagaaagag 282
      | || | ||| ||| || ||| ||||

```

Db 634 GAAAACTGCACCAGGTCTAAAATGGCACGGCTGAGTGAAGAG 592

RESULT 3

US-08-666-221B-11/c

; Sequence 11, Application US/08666221B

; Patent No. 6136544

; GENERAL INFORMATION:

; APPLICANT: Kamboj, Rajender

; APPLICANT: Nutt, Stephen

; TITLE OF INVENTION: GLUTAMATE RECEPTOR (OR EAA RECEPTOR)

; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USES

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/666,221B

; FILING DATE: 20-JUN-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Bent, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 016777/0308

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2878 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: sig_peptide

; LOCATION: 134..226

; FEATURE:

; NAME/KEY: mat_peptide

; LOCATION: 227..2860

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 134..2860

US-08-666-221B-11

Query Match 7.4%; Score 31; DB 3; Length 2878;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy      60 gattaaatgtcaacatttgccttttcgcgttccaattactaatgttacggcattattcag 119
      ||||| ||| | ||||| || | || | ||| ||| | | |||| ||
Db      814 GATTACATGAAACTCCTTGCCTCTTTTCATTTCTTTTAGTAAGGGTTTGCATCCTTTGT 755

Qy     120 gacagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgagag 179
      ||| | | |||| | || || | | | ||| ||| |||
Db      754 ATCAGCAGGTAAGTACGAATTTTGTGTCGAAGATTATACCTTGATGGAGCTTTGATGAG 695

Qy     180 tctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctga 239
      |||| | | || || || | || | ||| ||| |
Db      694 CTCTTGCAAACGAATGAGACCAGTGCTGTGCATCATACACAACCGTGACGGTTTTCCACTT 635

Qy     240 ctgggggtttcagcctatctgacatccaactgcagaagaaagag 282
      | || | || | || | || || | ||||
Db      634 GAAAAACTGCACCAGGTCTAAAATGGCACGGCTGAGTGAAGAG 592
```

RESULT 4

US-08-666-221B-13/c

; Sequence 13, Application US/08666221B

; Patent No. 6136544

; GENERAL INFORMATION:

; APPLICANT: Kamboj, Rajender

; APPLICANT: Nutt, Stephen

; TITLE OF INVENTION: GLUTAMATE RECEPTOR (OR EAA RECEPTOR)

; TITLE OF INVENTION: POLYNUCLEOTIDES AND THEIR USES

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/666,221B

; FILING DATE: 20-JUN-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Bent, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 016777/0308

; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2878 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 134..226
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 227..2860
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 134..2860
US-08-666-221B-13

```

```

Query Match          7.4%; Score 31; DB 3; Length 2878;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

```

Qy      60 gattaaatgtcaacatttgccttttcgcgttccaattactaatgttacggcattattcag 119
      ||||| ||| | ||||| || | || | ||| ||| | | |||| ||
Db      814 GATTACATGAAACTCCTTGCCTCTTTTCATTTCTTTTAGTAAGGGTTTTGCATCCTTTGT 755

Qy     120 gacagaactttactggaacgtcctgtgttcaatgcattctgtggaaaggaatgttgcagag 179
      ||| | | |||| | || || | | | | ||| ||| ||| |||
Db      754 ATCAGCAGGTAAGTACGAATTTTGAGTCGAAGATTATACCTTGATGGAGCTTTGATGAG 695

Qy     180 tctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctga 239
      |||| | | || || || | || | || | ||| | | |
Db      694 CTCTTGCAAACGAATGAGACCAGTGCTGTGCATCATACACAACCGTGACGGTTTTCCACTT 635

Qy     240 ctgggggtttcagcctatctgacatccaactgcagaagaaagag 282
      | || | ||| || | || || || ||||
Db      634 GAAAACTGCACCAGGTCTAAAATGGCACGGCTGAGTGAAGAG 592

```

```

RESULT 5
US-08-539-304A-5
; Sequence 5, Application US/08539304A
; Patent No. 5792933
; GENERAL INFORMATION:
; APPLICANT: MA, DIN-POW
; TITLE OF INVENTION: FIBER-SPECIFIC PROTEIN EXPRESSION IN THE
; TITLE OF INVENTION: COTTON PLANT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 JEFFERSON DAVIS HWY. SUITE 400
; CITY: ARLINGTON
; STATE: VA

```

```

; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/539,304A
; FILING DATE: 04-OCT-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN, OBLON F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2343-037-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(741..1093, 1220..1226)
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1094..1219
US-08-539-304A-5

```

```

Query Match          7.3%; Score 30.6; DB 1; Length 2646;
Best Local Similarity 60.0%; Pred. No. 1.5;
Matches 51; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

Qy 147 ttcaatgcattctgtgggaaaggaatgttgagagtcgtgtgcagccaaggagatgcaaggcc 206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 858 TGCATTGCTTACTTGAAAGGGAATGGTGCTGGTTCTGCTCCCCCAGCTTGCTGCAACGGC 917

Qy 207 atttgtggacgaagctgtgtgctgcaa 231
    | | | | | | | | | | | | | | | |
Db 918 ATCAGATCTCTCAACTCTGCCGCCA 942

```

```

RESULT 6
US-08-627-151A-5/c
; Sequence 5, Application US/08627151A
; Patent No. 5866341
; GENERAL INFORMATION:
; APPLICANT: SPINELLA, Dominic
; APPLICANT: BECHERER, Kathleen
; APPLICANT: BROWN, Steven
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: SCREENING DRUG LIBRARIES

```



```

;   NUMBER OF SEQUENCES:  19
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Gen-Probe Incorporated
;       STREET:    10210 Genetic Center Drive
;       CITY:      San Diego
;       STATE:     CA
;       COUNTRY:   USA
;       ZIP:       92121
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:     IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:     FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/627,151A
;       FILING DATE:       03-APR-1996
;       CLASSIFICATION:    435
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;       NAME:  Fisher, Carlos A
;       REGISTRATION NUMBER:  36,510
;       REFERENCE/DOCKET NUMBER:  CBI016
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-410-8926
;       TELEFAX:   619-410-8928
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  5:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH:  6253 base pairs
;           TYPE:    nucleic acid
;           STRANDEDNESS:  single
;           TOPOLOGY:  linear
US-08-627-151A-5

```

```

Query Match          7.3%;  Score 30.6;  DB 2;  Length 6253;
Best Local Similarity 53.8%;  Pred. No. 2.3;
Matches 63;  Conservative 0;  Mismatches 54;  Indels 0;  Gaps 0;

```

```

Qy   185 gcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactggg 244
      ||| | | | | | | | | | | | | | | | | | | | | |
Db   5791 GCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGATGCG 5732

Qy   245 gtttcagcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaac 301
      | | | | | | | | | | | | | | | | | | | | | |
Db   5731 GCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC 5675

```

```

RESULT 7
US-08-318-826A-9/c
; Sequence 9, Application US/08318826A
; Patent No. 5891725
; GENERAL INFORMATION:
;   APPLICANT:  Soreq, Hermona
;   APPLICANT:  Zakut, Haim

```

; APPLICANT: Eckstein, Fritz
; TITLE OF INVENTION: Synthetic Antisense
; TITLE OF INVENTION: Oligodeoxynucleotides and Pharmaceutical Compositions
; TITLE OF INVENTION: Containing Them
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5891725thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,826A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2391.00001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: fetal
; TISSUE TYPE: Brain, Liver
; POSITION IN GENOME:
; MAP POSITION: 3q26
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 160..1881
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /EC_number= 3.1.1.8
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "BCHE"
; OTHER INFORMATION: /note= "butyrylcholinesterase mature peptide"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 76..159
; FEATURE:
; NAME/KEY: mRNA
; LOCATION: 1..2381

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..1884
US-08-318-826A-9

Query Match 7.1%; Score 30; DB 2; Length 2381;
Best Local Similarity 50.5%; Pred. No. 2.2;
Matches 98; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

```
Qy 126 actttactggaacgtcctgtgttcaatgcattctgggaaaggaatggtgcagagtctgtg 185
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1485 AATTTTCATAGCCATGCATCACTCCCATCCATTCTGGCCACGGAAGTTTGGAGGATCGGTG 1426

Qy 186 cagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggg 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1425 TTCAAATAGTAGAAAAAGGCATTATTTCCCCATTCTGAG-AACTTCTTGGTGAAGTCCA 1367

Qy 246 tttcagcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1366 AGGCAGGGCATATGAAATTATAATCCCCAACACATCACCCAAGGCCTCACGGTAGTTTT 1307

Qy 306 cacgtctctgtttca 319
    | | | | | | | | | |
Db 1306 CAGGTCTCTGATCA 1293
```

RESULT 8
5215909-13/c
;Patent No. 5215909
; APPLICANT: SOREQ, HERMONA
; TITLE OF INVENTION: HUMAN CHOLINESTERASE GENES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/572,911
; FILING DATE: 15-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 87,724
; FILING DATE: 21-AUG-1987
; APPLICATION NUMBER: 875,737
; FILING DATE: 18-JUN-1986
;SEQ ID NO:13:
; LENGTH: 2400
5215909-13

Query Match 7.1%; Score 30; DB 6; Length 2400;
Best Local Similarity 50.5%; Pred. No. 2.2;
Matches 98; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

```
Qy 126 actttactggaacgtcctgtgttcaatgcattctgggaaaggaatggtgcagagtctgtg 185
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1485 AATTTTCATAGCCATGCATCACTCCCATCCATTCTGGCCACGGAAGTTTGGAGGATCGGTG 1426

Qy 186 cagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggg 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1425 TTCAAATAGTAGAAAAAGGCATTATTTCCCCATTCTGAG-AACTTCTTGGTGAAGTCCA 1367
```

Qy 246 tttcagcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
 ||| || ||| || | | | || | | ||| | | | |
 Db 1366 AGGCAGGGCATATGAAATTATAATCCCCAACAAACATCACCCAAGGCCTCACGGTAGTTTT 1307

Qy 306 cacgtctctgttca 319
 || ||||| |||
 Db 1306 CAGGTCTCTGATCA 1293

RESULT 9
 US-08-318-826A-8/c
 ; Sequence 8, Application US/08318826A
 ; Patent No. 5891725
 ; GENERAL INFORMATION:
 ; APPLICANT: Soreq, Hermona
 ; APPLICANT: Zakut, Haim
 ; APPLICANT: Eckstein, Fritz
 ; TITLE OF INVENTION: Synthetic Antisense
 ; TITLE OF INVENTION: Oligodeoxynucleotides and Pharmaceutical Compositions
 ; TITLE OF INVENTION: Containing Them
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kohn & Associates
 ; STREET: 30500 No. 5891725thwestern Hwy., Suite 410
 ; CITY: Farmington Hills
 ; STATE: Michigan
 ; COUNTRY: US
 ; ZIP: 48334
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/318,826A
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kohn, Kenneth I.
 ; REGISTRATION NUMBER: 30,955
 ; REFERENCE/DOCKET NUMBER: 2391.00001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (248) 539-5050
 ; TELEFAX: (248) 539-5055
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2416 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; TISSUE TYPE: Basal ganglion

```

; POSITION IN GENOME:
; MAP POSITION: 3q26
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 214..1935
; OTHER INFORMATION: /EC_number= 3.1.1.8
; OTHER INFORMATION: /gene= "BCHE"
; OTHER INFORMATION: /note= "butyrylcholinesterase mature peptide"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 130..213
; FEATURE:
; NAME/KEY: mRNA
; LOCATION: 1..2416
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 130..1938
US-08-318-826A-8

```

```

Query Match          7.1%; Score 30; DB 2; Length 2416;
Best Local Similarity 50.5%; Pred. No. 2.2;
Matches 98; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

```

```

Qy 126 actttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgacagagtctgtg 185
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1539 AATTTTCATAGCCATGCATCACTCCCATCCATTCTGGCCACGGAAGTTTGGAGGATCGGTG 1480

Qy 186 cagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggg 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1479 TTCAAATAGTAGAAAAAGGCATTATTTCCCCATTCTGAG-AACTTCTTGGTGAACCTCCA 1421

Qy 246 tttcagcctatctgacatccaactgcagaagaaagggtcaaggcttttttgaactcat 305
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1420 AGGCAGGGCATATGAAATTATAATCCCCAACACATCACCCAAGGCCTCACGGTAGTTTT 1361

Qy 306 cacgtctctgttca 319
    | | | | | | | | | |
Db 1360 CAGGTCTCTGATCA 1347

```

```

RESULT 10
US-09-334-489-1/c
; Sequence 1, Application US/09334489
; Patent No. 6291175
; GENERAL INFORMATION:
; APPLICANT: Pierre Seigny
; APPLICANT: Keith Schappert
; APPLICANT: Heiko Wiesbusch
; TITLE OF INVENTION: METHODS FOR TREATING A NEUROLOGICAL
; TITLE OF INVENTION: DISEASE BY DETERMINING BCHE GENOTYPE
; FILE REFERENCE: 08523/013002
; CURRENT APPLICATION NUMBER: US/09/334,489
; CURRENT FILING DATE: 1999-06-16
; PRIOR APPLICATION NUMBER: 60/089,406
; PRIOR FILING DATE: 1998-06-18
; NUMBER OF SEQ ID NOS: 8

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-334-489-1

Query Match 7.1%; Score 30; DB 4; Length 2416;
Best Local Similarity 50.5%; Pred. No. 2.2;
Matches 98; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

```
Qy 126 actttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgcaagagtctgtg 185
    | ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1539 AATTTTCATAGCCATGCATCACTCCCATCCATTCTGGCCACGGAAGTTTGGAGGATCGGTG 1480

Qy 186 cagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggg 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1479 TTCAAATAGTAGAAAAAGGCATTATTTCCCCATTCTGAG-AACTTCTTGGTGAAGTCCA 1421

Qy 246 tttcagcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
    ||| | | ||| | | | | | | | | | | | | | | | | | | | | |
Db 1420 AGGCAGGGCATATGAAATTATAATCCCCAACACATCACCCAAGGCCTCACGGTAGTTTT 1361

Qy 306 cagctctctgttca 319
    || ||||| |||
Db 1360 CAGGTCTCTGATCA 1347
```

RESULT 11
US-09-334-489-2/c
; Sequence 2, Application US/09334489
; Patent No. 6291175
; GENERAL INFORMATION:
; APPLICANT: Pierre Sevigny
; APPLICANT: Keith Schappert
; APPLICANT: Heiko Wiesbusch
; TITLE OF INVENTION: METHODS FOR TREATING A NEUROLOGICAL
; TITLE OF INVENTION: DISEASE BY DETERMINING BCHE GENOTYPE
; FILE REFERENCE: 08523/013002
; CURRENT APPLICATION NUMBER: US/09/334,489
; CURRENT FILING DATE: 1999-06-16
; PRIOR APPLICATION NUMBER: 60/089,406
; PRIOR FILING DATE: 1998-06-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-334-489-2

Query Match 7.1%; Score 30; DB 4; Length 2416;
Best Local Similarity 50.5%; Pred. No. 2.2;
Matches 98; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

Qy 126 actttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgacagagtctgtg 185
 | ||| | | | | | | | | | | | | | | | | | |
 Db 1539 AATTTTCATAGCCATGCATCACTCCCATCCATTCTGGCCACGGAAGTTTGGAGGATCGGTG 1480

Qy 186 cagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggg 245
 || | | | | | | | | | | | | | | | | | |
 Db 1479 TTCAAATAGTAGAAAAAGGCATTATTTCCCCATTCTGAG-AACTTCTTGGTGAACCTCA 1421

Qy 246 tttcagcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaactcat 305
 ||| || ||| || | | | | | | | | | | | | | |
 Db 1420 AGGCAGGGCATATGAAATTATAATCCCCAACATCACCCAAGGCCTCACGGTAGTTTT 1361

Qy 306 cacgtctctgttca 319
 || ||||| |||
 Db 1360 CAGGTCTCTGATCA 1347

RESULT 12
 US-08-221-817-12
 ; Sequence 12, Application US/08221817
 ; Patent No. 5532151
 ; GENERAL INFORMATION:
 ; APPLICANT: Chantry, David
 ; APPLICANT: Gray, Patrick W.
 ; APPLICANT: Hoekstra, Merle F.
 ; TITLE OF INVENTION: A No. 5532151el G Protein-Coupled Receptor
 ; TITLE OF INVENTION: Kinase GRK6
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
 ; ADDRESSEE: Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/221,817
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/123,932
 ; FILING DATE: 17 SEP 1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 5532151and, Greta E.
 ; REGISTRATION NUMBER: 35,302
 ; REFERENCE/DOCKET NUMBER: 31981
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (312) 474-6300
 ; TELEFAX: (312) 474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2204 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1758
US-08-221-817-12

Query Match 6.9%; Score 29; DB 1; Length 2204;
Best Local Similarity 61.0%; Pred. No. 4.7;
Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 122 cagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgcagagtc 181
|||| || ||||| || || | | || || ||||| ||
Db 1634 CAGACCTGGACTGGAAGGGCCAGCCACCTGCACCTCCTAAAAAGGGACTGCTGCAGAGAC 1693

Qy 182 tgtgcagccaaggagat 198
| | ||| | ||||
Db 1694 TCTTCAGTCGCCAAGAT 1710

RESULT 13

US-08-454-439-12

; Sequence 12, Application US/08454439

; Patent No. 5591618

; GENERAL INFORMATION:

; APPLICANT: Chantry, David

; APPLICANT: Gray, Patrick W.

; APPLICANT: Hoekstra, Merle F.

; TITLE OF INVENTION: A No. 5591618el G Protein-Coupled Receptor

; TITLE OF INVENTION: Kinase GRK6

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Borun

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/454,439

; FILING DATE: 30-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/221,817

; FILING DATE: 31-MAR-1994

; APPLICATION NUMBER: 08/123,932


```

; FILING DATE: 17 SEP 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5591618and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2204 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1758
US-08-454-439-12

```

```

Query Match          6.9%; Score 29; DB 1; Length 2204;
Best Local Similarity 61.0%; Pred. No. 4.7;
Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```

```

Qy 122 cagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgagagtc 181
      |||| || ||||| | || | | || || || ||||| |
Db 1634 CAGACCTGGACTGGAAGGGCCAGCCACCTGCACCTCCTAAAAAGGGACTGCTGCAGAGAC 1693

Qy 182 tgtgcagccaaggagat 198
      || ||| | |||
Db 1694 TCTTCAGTCGCCAAGAT 1710

```

```

RESULT 14
PCT-US94-10487-12
; Sequence 12, Application PC/TUS9410487
; GENERAL INFORMATION:
; APPLICANT: ICOS Corporation
; TITLE OF INVENTION: A Novel G Protein-Coupled Receptor
; TITLE OF INVENTION: Kinase GRK6
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10487
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,817
; FILING DATE: 31 MAR 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2204 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..1758
PCT-US94-10487-12

```

```

Query Match          6.9%; Score 29; DB 5; Length 2204;
Best Local Similarity 61.0%; Pred. No. 4.7;
Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```

```

Qy 122 cagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgcagagtc 181
    ||| || ||||| | || | | || || || ||||| |
Db 1634 CAGACCTGGACTGGAAGGGCCAGCCACCTGCACCTCCTAAAAAGGGACTGCTGCAGAGAC 1693

Qy 182 tgtgcagccaaggagat 198
    | ||| | |||
Db 1694 TCTTCAGTCGCCAAGAT 1710

```

```

RESULT 15
US-08-221-817-10
; Sequence 10, Application US/08221817
; Patent No. 5532151
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5532151el G Protein-Coupled Receptor
; TITLE OF INVENTION: Kinase GRK6
; NUMBER OF SEQUENCES: 24

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 08:20:37 ; Search time 4942.22 Seconds
(without alignments)
915.373 Million cell updates/sec

Title: US-09-394-745-5950
Perfect score: 421
Sequence: 1 ggggtccaggcacgcgtccga.....agtggcagaatttggtgccgc 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%		Query		DB	ID	Description
	No.	Score	Match	Length			
	1	193.6	46.0	578	13	AQ330565	AQ330565 nbxb0047D
	2	146.6	34.8	742	10	AW448782	AW448782 BRY_1421
	3	94.8	22.5	462	11	BG238542	BG238542 sab48e05.
	4	73.8	17.5	408	10	AW030046	AW030046 EST273301
	5	63.6	15.1	658	11	BI271227	BI271227 NF051E08F
	6	56.2	13.3	595	11	BG606545	BG606545 WHE2957_G
	7	54.6	13.0	643	11	BI271547	BI271547 NF057E12F
c	8	54.2	12.9	595	11	BG321058	BG321058 Zm04_01a1
	9	47.6	11.3	479	11	BG507333	BG507333 sac57f11.
	10	44	10.5	518	10	AL372617	AL372617 MtBA52C12
	11	39.4	9.4	414	13	AZ406647	AZ406647 1M0175P21
	12	39	9.3	370	13	AQ083972	AQ083972 HS_2226_B
	13	36.4	8.6	584	13	AZ873903	AZ873903 2M0187P15
c	14	36	8.6	596	10	AW280435	AW280435 fj40d12.y
c	15	35.4	8.4	498	13	AZ637668	AZ637668 1M0497K12
c	16	34.8	8.3	300	11	C55571	C55571 C55571 Yuji
	17	34.8	8.3	909	13	AZ675997	AZ675997 ENTMU59TR
	18	34.6	8.2	919	11	BF689537	BF689537 602186918
c	19	34.4	8.2	321	10	BB143803	BB143803 BB143803
c	20	34.4	8.2	806	10	AL580988	AL580988 AL580988
	21	34.2	8.1	500	11	W53261	W53261 md40f02.r1
	22	34.2	8.1	629	11	C23806	C23806 C23806 Dict
c	23	34.2	8.1	786	10	AU135670	AU135670 AU135670
c	24	34	8.1	478	13	AZ035699	AZ035699 RPCI-23-3
	25	34	8.1	771	11	BF864675	BF864675 963053G04
c	26	33.8	8.0	1003	13	CNS03FGK	AL241661 Tetraodon
	27	33.6	8.0	726	13	AQ377821	AQ377821 RPCI11-16
	28	33.4	7.9	250	13	B82721	B82721 RPCI11-17I1
	29	33.4	7.9	691	11	BG540863	BG540863 602570011
	30	33.2	7.9	413	10	AA234489	AA234489 zr74h03.r
	31	33.2	7.9	509	10	AW450334	AW450334 UI-H-BI3-
c	32	33.2	7.9	514	13	AQ149197	AQ149197 HS_3164_B
c	33	33.2	7.9	566	13	BH002069	BH002069 BMBAC02A0
	34	33.2	7.9	717	13	AZ728677	AZ728677 RPCI-24-7
	35	33.2	7.9	920	13	AZ540221	AZ540221 ENTGJ45TR
c	36	32.8	7.8	346	10	AA318887	AA318887 EST21401
	37	32.8	7.8	460	10	AA412635	AA412635 zt97c12.r
c	38	32.8	7.8	573	13	AZ228922	AZ228922 RPCI-23-6
	39	32.8	7.8	594	13	AZ553701	AZ553701 RPCI-23-1
	40	32.8	7.8	615	13	AQ918240	AQ918240 RPCI-23-2
	41	32.8	7.8	735	13	AZ374574	AZ374574 1M0127K24
	42	32.8	7.8	1025	11	BG872410	BG872410 602792712
c	43	32.6	7.7	432	13	AQ584680	AQ584680 RPCI-11-4
c	44	32.6	7.7	573	10	BE204987	BE204987 EST397663
	45	32.6	7.7	702	10	AL135286	AL135286 DKFZp762D

ALIGNMENTS

RESULT 1

AQ330565

LOCUS

AQ330565

578 bp

DNA

GSS

08-JAN-1999

DEFINITION nbxb0047D17f CUGI Rice BAC Library Oryza sativa genomic clone
 nbxb0047D17f, DNA sequence.

ACCESSION AQ330565

VERSION AQ330565.1 GI:4122415

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 578)

AUTHORS Wing,R.A. and Dean,R.A.

TITLE A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL Unpublished (1998)

COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TAATACGACTCACTATAGGG
 Class: BAC ends
 High quality sequence stop: 292.

FEATURES

source Location/Qualifiers

1. .578
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nbxb0047D17f"
 /clone_lib="CUGI Rice BAC Library"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
 HindIII; Rice is one of two most popular grains in the
 world. Half of the world population especially those
 inhabiting highly populated areas of the humid tropics
 and subtropics, rely on rice as their primary source of
 carbohydrate. Monocotyledonous rice is a diploid plant
 (2n=24) with a haploid genome equivalent of 431 Mbp
 (Arumuganathan and Earle, 1991). The relatively small
 genome of rice, three times larger than that of
 Arabidopsis, makes it suitable for genomic studies. In
 order to facilitate positional cloning, physical mapping
 and genome sequencing of rice, we have constructed a BAC
 library from Oryza sativa, Nipponbare variety. The
 library contains 36,864 clones with an average insert size
 of 128.5 Kb providing 10.9 haploid genome equivalents. The
 deep coverage allows the isolation a particular sequence
 with a probability of 99.9 %. Two high density filters,
 each containing 18,432 clones (doubly spotted), represent
 the whole library for colony screening."

BASE COUNT 152 a 114 c 111 g 201 t

ORIGIN

Query Match 46.0%; Score 193.6; DB 13; Length 578;
 Best Local Similarity 81.6%; Pred. No. 3.8e-47;
 Matches 248; Conservative 0; Mismatches 54; Indels 2; Gaps 2;

```

Qy      71 aacatttgcccttttcgcggttccaattactaatgttacg-gcattattcaggacagaactt 129
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      25 AGCATTGACCCTTCATCATGCCAAATAATAATCTTATGTGCATTACTCAGGACAAAACCTT 84

Qy     130 tactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgagagtcgtgcagc 189
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      85 CACTGGAAAGTCCTATGTTCAATGCATTTTGGGAAAAGGATGTTGCAGAGTCTGTGC-GC 143

Qy     190 caaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggggtttc 249
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     144 CAAGGAGATGCACAGCCTTTTGTAGAGGAAGCTGTACTGCAAGTATCAGATTGGGGATTTC 203

Qy     250 agcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaactcatcacg 309
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     204 AGCTTGTCAGACATTCAAATGCAGAAGAGAGAGGATCTGAGCTTTTTTGAATTGATCAAA 263

Qy     310 tctctgttcaatcatgctgaaaaacagtggggtgggatttctggggccaatacatatatcg 369
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     264 TCTCTATTCCGTCAGGCTGAACGGGAGTGGGTGGGATTTCTGGGCCCAATACACATATGG 323

Qy     370 cagg 373
      | | | |
Db     324 CAGG 327
  
```

RESULT 2

AW448782

LOCUS AW448782 742 bp mRNA EST 03-JAN-2001
 DEFINITION BRY_1421 BRY Triticum aestivum cDNA clone P35-10, mRNA sequence.
 ACCESSION AW448782
 VERSION AW448782.1 GI:12019317
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 742)
 AUTHORS Clarke,B.C., Hobbs,M. and Appels,R.
 TITLE Genes active in developing wheat endosperm
 JOURNAL Unpublished (2000)
 COMMENT Contact: Bryan Clarke
 Division of Plant Industry
 C.S.I.R.O.
 GPO Box 1600, Canberra, ACT, Australia
 Tel: 61 2 6246 5054
 Fax: 61 2 6246 5000
 Email: bryanc@pi.csiro.au.

FEATURES Location/Qualifiers
 source 1. .742
 /organism="Triticum aestivum"
 /cultivar="Wyuna"
 /db_xref="taxon:4565"

```

        /clone="P35-10"
        /clone_lib="BRY"
        /cell_type="endosperm"
BASE COUNT      180 a      172 c      218 g      169 t      3 others
ORIGIN

```

```

Query Match          34.8%;  Score 146.6;  DB 10;  Length 742;
Best Local Similarity 76.8%;  Pred. No. 4.1e-33;
Matches 179;  Conservative 0;  Mismatches 54;  Indels 0;  Gaps 0;

```

```

Qy   188  gccaaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggggtt 247
      |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db    3  GCCCCGGGAGATGCGCAACCATTTGTAGAGGAAGCTGTGCTGCACGTATCTGATTGGGGAT 62

Qy   248  tcagcctatctgacatccaactgcagaagaaagaggctcaaggcttttttgaactcatca 307
      ||||  ||  |||||  ||  |||||  |||||  |||||  ||||  ||  |||||  |||||
Db    63  TCAGTTTGTGACACATTACATGCAGAAGAAAGAGGATCAGGGAGTATTTGAATTTATCA 122

Qy   308  cgtctctgttcaatcatgctgaaaaacagtgggtgggatttctgggccaatacatatat 367
      |||||  |||  ||  |||||  |  |||||  |||||  |||||  |||||  ||  ||  |
Db   123  AGTCTCTGATCAGTCAGGCTGAACGAGAGTGGGTGGGATTTCTGGTCCCAATCCACATCT 182

Qy   368  cgcaggggatagatgaccgagtgatctcgccctcagtggcagaatttgtgccg 420
      |  ||||  ||  |||||  |||  |  |  ||  ||  ||  ||  ||||  ||  |
Db   183  GGTAGGGAATGGATGACCGGGTGGTGCCCCCATCGGCGACCGAGTTTGCCCGG 235

```

```

RESULT      3
BG238542
LOCUS       BG238542      462 bp      mRNA          EST          13-FEB-2001
DEFINITION  sab48e05.y1 Gm-cl043' Glycine max cDNA clone GENOME SYSTEMS CLONE
              ID: Gm-cl043-2529 5' similar to TR:Q9SF34 Q9SF34 F11F8.28 PROTEIN.
              ;; mRNA sequence.
ACCESSION   BG238542
VERSION     BG238542.1  GI:12773615
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
              Glycine.
REFERENCE   1  (bases 1 to 462)
AUTHORS     Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
              ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
              Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
              ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
              ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
              ,R., Waterston,R. and Wilson,R.
TITLE       Public Soybean EST Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
              Public Soybean EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800

```


Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 397.

FEATURES Location/Qualifiers
 source 1. .462
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1043-2529"
 /clone_lib="Gm-c1043"
 /tissue_type="Hypocotyl and Plumule, germinating seeds"
 /lab_host="DH10B"
 /note="Vector: pT7T3Pac (Pharmacia); Site_1: EcoRI;
 Site_2: NotI; This cDNA library was constructed from mRNA
 isolated from hypocotyl and plumule tissues of seeds
 germinated for three days of the cultivar Williams.
 Complementary DNA was synthesized from mRNA using a primer
 consisting of a poly(dT) sequence with a NotI restriction
 site. EcoRI adapters were ligated to the blunt-ended cDNA
 fragments followed by digestion with EcoRI and NotI. The
 cDNA fragments were directionally cloned into the
 EcoRI-NotI restriction site of the pT7T3-Pac vector. The
 ligated cDNA fragments were transformed into DH10B host
 cells (Gibco BRL). This library was constructed by Dr.
 Randy Shoemaker."
 BASE COUNT 133 a 87 c 115 g 127 t
 ORIGIN

Query Match 22.5%; Score 94.8; DB 11; Length 462;
 Best Local Similarity 61.6%; Pred. No. 1e-17;
 Matches 186; Conservative 0; Mismatches 112; Indels 4; Gaps 2;

```

Qy   118  aggacagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgag 177
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   98  AGGATAAACTTATGATTGAAGAACCAGAATTTGAGGAATTTGGCAGAGGGATGTGGAGG 157

Qy   178  agtctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatct 237
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   158  AGTCAGTTC-GTCAGGGAAACATACGCCCATTTATAGAAGAAGCTGTTCTGCAGGTATCA 216

Qy   238  gactgggggtttcagcctatctgacatccaactgcagaagaa---agaggctcaaggcttt 294
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   217  AATTGGGGTTTTGACCTTAAGGAACCTCATGTGCAAAAGAAGTGTCAAACAAGAGGCATA 276

Qy   295  tttgaactcatcacgtctctgttcaatcatgctgaaaaacagtgggtgggatttctgggc 354
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   277  CTTCTTTGGTTGAAATCCATGTACAGTCAGGCGGACTGTGAATTAGCAGGATTTCTTGGC 336

Qy   355  ccaatacatatatcgcaggggatagatgaccgagtgatctcgccctcagtggcagaattt 414
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   337  CTTACACATATATGGCAGGGACTGGATGATAGGGTGGTCCCACCATCAGTGATGGAATAT 396

```

Qy 415 gt 416
|
Db 397 AT 398

RESULT 4

AW030046

LOCUS AW030046 408 bp mRNA EST 18-MAY-2001
DEFINITION EST273301 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEC16N4, mRNA sequence.

ACCESSION AW030046

VERSION AW030046.1 GI:5888802

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 408)

AUTHORS Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Ronning,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.

TITLE Generation of ESTs from tomato callus tissue

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

Clemson University Genomics Institute
Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES Location/Qualifiers

source

1. .408

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cLEC16N4"

/clone_lib="tomato callus, TAMU"

/tissue_type="callus"

/dev_stage="25-40 days old"

/lab_host="XL1-Blue MRF"

/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 122 a 85 c 87 g 114 t

ORIGIN

Query Match 17.5%; Score 73.8; DB 10; Length 408;

Best Local Similarity 59.9%; Pred. No. 1.8e-11;

Matches 142; Conservative 0; Mismatches 92; Indels 3; Gaps 1;

Qy 188 gccaaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatctgactgggggtt 247
| ||| ||||| ||||| || ||||| | || || || || ||||| |

Db	5	GACAAAAGAATGCAAAACCATTTGTAGAGGAAGCTGTCTTACAGGTTTCCAATTGGGGAT	64
Qy	248	tcagcctatctgacatccaactgcagaagaaagaggctc---aaggcttttttgaactca	304
Db	65	TTAGTCCTGCAGACCTCAAAGTACAGAGGACACGCACTGGGAAGGGTATTATGCATTGGA	124
Qy	305	tcacgtctctgttcaatcatgctgaaaaacagtgggtgggatttctgggccaatacata	364
Db	125	TTAAATCTCTATTTGGTCAAACAGACGAAATCTTGACTGGATTCTTGGTCAAATACATG	184
Qy	365	tatcgaggggatagatgaccgagtgatctcgccctcagtggcagaatttgtgccgc	421
Db	185	TATGGCAGGGAATGGAAGATATGGTGGTACCGCCATCCACAAGTGATTTCTTGCAGC	241

RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-ZAP XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

BASE COUNT 191 a 122 c 160 g 183 t 2 others
ORIGIN

Query Match 15.1%; Score 63.6; DB 11; Length 658;
Best Local Similarity 58.3%; Pred. No. 2.2e-08;
Matches 147; Conservative 0; Mismatches 101; Indels 4; Gaps 2;

```

Qy 118 aggacagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgag 177
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 408 AGGATGAAATTCTCGTCGATGAACCAGCATTCTGAAGAGTATTGGCAGAGGGATCTGGAGG 467

Qy 178 agtctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatct 237
    |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 468 AGTCTGTTC-GGCAGGGAAACCTGAAGCCGTTTATAGAGGAAGCTCTTCTGCAGGTATCT 526

Qy 238 gactgggggtttcagcctatctgacatccaactgcagaagaa---agaggctcaaggcttt 294
    |||  ||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 527 AGATGGGATTTCAACATANAAGAACTTCATGTGCATAAGAAGTGTCAAACAGGAGGATTA 586

Qy 295 tttgaactcatcacgtctctgttcaatcatgctgaaaaacagtgggtgggatttctgggc 354
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 587 CTTCTTTGGTTGAAATCCATGTACGGTCAGGCAGAATGTGAATTANCAGGATATCTCGGC 646

Qy 355 ccaatacatata 366
    |  ||||  |||
Db 647 CGTATACACATA 658

```

RESULT 6
BG606545
LOCUS BG606545 595 bp mRNA EST 17-APR-2001
DEFINITION WHE2957_G02_N03ZS Wheat dormant embryo cDNA library Triticum aestivum cDNA clone WHE2957_G02_N03, mRNA sequence.
ACCESSION BG606545
VERSION BG606545.1 GI:13656528
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 595)
AUTHORS Anderson,O.D., Chao,S., Chin,A., Close,T.J., Doherty,L., Fenton ,R.D., Lazo,G.R., Rausch,C.J., Walker-Simmons,M.K. and Wilson,C.
TITLE The structure and function of the expressed portion of the wheat genomes - Dormant embryo cDNA library
JOURNAL Unpublished (2001)

COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oandersn@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.

FEATURES Location/Qualifiers
 source 1. .595
 /organism="Triticum aestivum"
 /cultivar="Brevor"
 /db_xref="taxon:4565"
 /clone="WHE2957_G02_N03"
 /clone_lib="Wheat dormant embryo cDNA library"
 /tissue_type="Seed embryo"
 /dev_stage="Mature seed"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown to seed
 maturity under conditions favoring seed dormancy (L.
 Doherty at K. Walker_Simmons lab, Washington State
 University, Pullman, WA). Embryos were cut from mature
 dormant seed (Doherty). Total RNA was prepared from these
 embryos, polyA was purified, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the TJ Close lab at the University of
 California, Riverside (Chin, Fenton). Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."

BASE COUNT 149 a 131 c 148 g 166 t 1 others
ORIGIN

Query Match 13.3%; Score 56.2; DB 11; Length 595;
Best Local Similarity 88.4%; Pred. No. 3.5e-06;
Matches 61; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 118 aggacagaactttactggaacgtcctgtgttcaatgcattctgggaaaggaatgttgcag 177
 ||||| ||| ||||||||| |||| ||||||| ||||||||||| | |||||||||
Db 525 AGGACAAAACCTTACTGGAAGCTCCTATGTTCAACGCATTCTGGGAAAAGGATGTTGCAG 584

Qy 178 agtctgtgc 186
 |||||||
Db 585 AGTCTGTGC 593

RESULT 7
BI271547
LOCUS BI271547 643 bp mRNA EST 18-JUL-2001
DEFINITION NF057E12FL1F1099 Developing flower Medicago truncatula cDNA clone
 NF057E12FL 5', mRNA sequence.
ACCESSION BI271547
VERSION BI271547.1 GI:14880151
KEYWORDS EST.

Db 468 AGTCTGTTC-GGCAGGGAAACCTGAAGCCGTTTATAGAGGAAGCTCTTCTGCAGGTATCT 526

Qy 238 gactgggggtttcagcctatctgacatccaactgcagaagaa 278
 |||| |||| | || | || |||| ||||

Db 527 AGATGGGATTTCAACATAGAAGAACTTCATGTGCATAAGAA 567

RESULT 8
 BG321058/c
 LOCUS BG321058 595 bp mRNA EST 27-FEB-2001
 DEFINITION Zm04_01a12_A Zm04_AAFC_ECORC_cold_stressed_maize_seedlings Zea mays
 cDNA clone Zm04_01a12, mRNA sequence.
 ACCESSION BG321058
 VERSION BG321058.1 GI:13150736
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 595)
 AUTHORS Singh,J.A., Wakui,K., Couroux,P., De Moors,A., Harris,L.J., Hattori
 ,J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A.
 TITLE Expressed Sequence Tags from Cold-Stressed Maize Seedlings
 JOURNAL Unpublished (2001)
 COMMENT Contact: Singh,J.A.
 Eastern Cereal and Oilseed Research Centre
 Agriculture and Agri-food Canada
 960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada
 Tel: (613) 759-1662
 Fax: (613) 759-1701
 Email: singhja@em.agr.ca.
 FEATURES Location/Qualifiers
 source 1: .595
 /organism="Zea mays"
 /cultivar="CO328"
 /db_xref="taxon:4577"
 /clone="Zm04_01a12"
 /clone_lib="Zm04_AAFC_ECORC_cold_stressed_maize_seedlings"
 /tissue_type="Leaf, crown"
 /note="Vector: Bluescript SK-/XhoI-EcoRI; Site_1: Eco RI;
 Site_2: Xho I; Lower temperature 5o C / hour from 22 to
 12oC; bring to 5o in 1 hour from 12oC. Leave at 5oC 2 days
 , photoperiod 16 hours. Light intensity was 125 uE-1.
 Library prepared by in vivo mass excision from amplified
 library."
 BASE COUNT 134 a 149 c 156 g 149 t 7 others
 ORIGIN

Query Match 12.9%; Score 54.2; DB 11; Length 595;
 Best Local Similarity 88.1%; Pred. No. 1.4e-05;
 Matches 59; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 355 ccaatacatatatatcgaggggatagatgaccgagtgatctcgccctcagtggcagaat 414
 ||||| ||||| || ||||| ||||| ||||| |||||

Db 595 CCAATACATATATGGCAGGGGATGGACGACCGAGTGGTCTCGCCGGCAGTGGCCGAATTT 536

Qy 415 gtgccgc 421
 |||| ||
Db 535 GTGCGGC 529

RESULT 9

BG507333

LOCUS BG507333 479 bp mRNA EST 28-MAR-2001

DEFINITION sac57f11.y1 Gm-cl062 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl062-4125 5' similar to TR:Q9SF34 Q9SF34 F11F8.28 PROTEIN.
;, mRNA sequence.

ACCESSION BG507333

VERSION BG507333.1 GI:13477451

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE 1 (bases 1 to 479)

AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or

info@genomesystems.com web site: www.genomesystems.com

High quality sequence stop: 438.

FEATURES

source

Location/Qualifiers

1. .479

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl062-4125"

/clone_lib="Gm-cl062"

/tissue_type="stem tissue of greenhouse grown plants"

/dev_stage="1 month old"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from stem tissue of 1 month old greenhouse grown plants
for the cultivar Raiden. Complementary DNA was
synthesized from mRNA using a primer consisting of a

/db_xref="taxon:3880"
 /clone="MtBA52C12"
 /clone_lib="MtBA"
 /tissue_type="root tips"
 /dev_stage="harvested after 3 days of N-starvation"
 /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
 XhoI; Plants were grown in an aeroponic chamber for 14
 days on nitrogen-rich medium followed by 3 days on N-free
 medium. RNA was extracted from root tips (1-3 cm). cDNA
 was prepared from polyA+ enriched RNA. The cDNA was
 directionally ligated into Uni-zapXR vector from
 Stratagene and packaged using Gigapack Gold packaging
 extracts. Plasmids containing cDNA inserts were
 mass-excised from phage stocks using ExAssit helper phage
 and propagated in SOLR cells. Clone ordering and
 sequencing was performed by the Centre National de
 Sequencage (Genoscope, Evry, France)."

BASE COUNT 174 a 112 c 95 g 137 t
 ORIGIN

Query Match 10.5%; Score 44; DB 10; Length 518;
 Best Local Similarity 73.7%; Pred. No. 0.015;
 Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 343 ggattttctggggcccaatacatatatcgcgaggggatagatgaccgagtgatctcgccctca 402
 || ||||| || ||||| |||| ||| || || ||| |||| | || ||||
 Db 19 GGCTTTCTCGGTCCAATACACATATGGCAAGGAATGGACGACAAAGTGGTTCCCCCATCA 78

 Qy 403 gtggcagaatttgtgc 418
 || | || |||||
 Db 79 ATGACTGATTTTGTGC 94

RESULT 11
 AZ406647
 LOCUS AZ406647 414 bp DNA GSS 03-OCT-2000
 DEFINITION 1M0175P21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0175P21 R, DNA sequence.
 ACCESSION AZ406647
 VERSION AZ406647.1 GI:10530660
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 414)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D.,Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah

LOCUS AQ083972 370 bp DNA GSS 26-AUG-1998
 DEFINITION HS_2226_B2_E10_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2226 Col=20 Row=J, DNA sequence.
 ACCESSION AQ083972
 VERSION AQ083972.1 GI:3452889
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 370)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2226 row: J column: 20
 Class: BAC ends
 High quality sequence stop: 370.
 FEATURES Location/Qualifiers
 source 1. .370
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2226 Col=20 Row=J"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 110 a 100 c 65 g 95 t
 ORIGIN

Query Match 9.3%; Score 39; DB 13; Length 370;
 Best Local Similarity 65.5%; Pred. No. 0.42;
 Matches 57; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 36 acaattccttagtagtcaccccttcgattaaatgtcaacatttgccttttcgcgttccaat 95
 | ||| || | | |||| || | | | || ||||| ||| ||||| ||| ||
 Db 250 AGAACTCCCAATTGTCAGAGCTACAAATCCCTGGCAACATAGCCCTGTTTCGCATTACAT 309
 Qy 96 tactaatgttacggcattattcaggac 122
 | ||||| ||| | || ||| || || |
 Db 310 TCCTAATGTTCTGCCTTAATCCGGCC 336

RESULT 13
 AZ873903

LOCUS AZ873903 584 bp DNA GSS 21-FEB-2001
 DEFINITION 2M0187P15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0187P15 R, DNA sequence.
 ACCESSION AZ873903
 VERSION AZ873903.1 GI:13082429
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 584)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0187 row: P column: 15
 Seq primer: CACACAGGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 584.
 FEATURES
 source Location/Qualifiers
 1. .584
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0187P15"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 144 a 92 c 182 g 166 t
ORIGIN

Query Match 8.6%; Score 36.4; DB 13; Length 584;
Best Local Similarity 54.5%; Pred. No. 2.8;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy      94 attactaatgttacggcattattcaggacagaactttactggaacgtcctgtgttcaatg 153
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     429 AGTGATCATGGATGTGCATATGTCTGTGCAGCAAGGAACATGGATGTGCATATGTCTGTG 488

Qy     154 cattctgggaaaggaatgttgcagagtctgtgcagccaaggagatgcaaggccatttgtg 213
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     489 CAGCAAGGAACATGGATGTGCATGTGTCTGTGCAGCAGGGAACATGGATGTGCATGTGTC 548

Qy     214 gacgaagctgtgct 227
      | | | | | | | |
Db     549 TATGTAGCAGTGAT 562
```

RESULT 14

AW280435/c

LOCUS AW280435 596 bp mRNA EST 04-JAN-2000
DEFINITION fj40d12.y1 zebrafish adult brain Danio rerio cDNA 5' similar to
SW:AP19_HUMAN P56377 CLATHRIN COAT ASSEMBLY PROTEIN AP19 ;, mRNA
sequence.

ACCESSION AW280435

VERSION AW280435.1 GI:6668984

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasborinae; Danio.

REFERENCE 1 (bases 1 to 596)

AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
 ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
 ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
 Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
 and Wilson,R.

TITLE WashU Zebrafish EST Project 1998

JOURNAL Unpublished (1998)

COMMENT Other_ESTs: fj40d12.x1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: John Ngai. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)

Seq primer: T3 ET from Amersham
High quality sequence stop: 488.

FEATURES
 source
 Location/Qualifiers
 1. .596
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish adult brain"
 /sex="mixed male and female"
 /tissue_type="brain"
 /dev_stage="adult"
 /lab_host="E. coli DH10B"
 /note="Vector: pZIPLOX; Site_1: NotI; Site_2: SalI;
 Original library was constructed in lambdaZIPLOX. Mass
 excision of the cDNA library was performed to yield
 pZIPLOX plasmids. Insert check was done in original
 library."

BASE COUNT 154 a 163 c 164 g 115 t
ORIGIN

Query Match 8.6%; Score 36; DB 10; Length 596;
Best Local Similarity 60.0%; Pred. No. 3.7;
Matches 60; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 177 gagtctgtgcagccaaggagatgcaaggccatttgtggacgaagctgtgctgcaagtatc 236
 ||||| ||||| | || || ||| | || | | | | | | | |
Db 556 GAGTGTGTGCAAGCGAGTGTGTGTGAGGTCTTCATGTGAGTCCGATCTCCTCCAGGACGC 497

Qy 237 tgactgggggtttcagcctatctgacatccaactgcagaag 276
 || ||||| ||||| || | | | | | | | | ||||| ||
Db 496 TGCGTGGGGTTTCTGCTTCTTTGGCGTCCTCCTGCAGTAG 457

RESULT 15
AZ637668/c
LOCUS AZ637668 498 bp DNA GSS 13-DEC-2000
DEFINITION 1M0497K12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0497K12 F, DNA sequence.
ACCESSION AZ637668
VERSION AZ637668.1 GI:11759858
KEYWORDS GSS.
SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 498)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D.,Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0497 row: K column: 12
 Seq primer: CGTTGTAAAACGACGGCCAGT
 Class: plasmid ends
 High quality sequence stop: 498.

FEATURES Location/Qualifiers
 source 1. .498
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0497K12"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptored DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptored mouse DNA was annealed to
 adaptored vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 169 a 107 c 81 g 141 t
 ORIGIN

Query Match 8.4%; Score 35.4; DB 13; Length 498;
 Best Local Similarity 50.9%; Pred. No. 5.4;
 Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy      23  tgagggttagcttaacaattcttagtagtcaccccttcgattaaatgtcaacatttgcctt 82
      |||  ||| ||  |  |  ||||  || |||||  ||  ||  |  ||  |||
Db      223  TGAACCTAGTTTTCTCGTGTTCTGTAGGAACGCCTTCTAATAGTTGATCTGAGTTTTCTT 164

Qy      83  ttcgcggttccaattactaatgttacggcattattcaggacagaactttactggaacgtcc 142
      ||  || |  ||  |  |||  |  |  |||  |  || |  |  |||  ||
Db      163  TTATTGTGCTGTTTTACACTGTGATAATAAAATGCAGCAATTCACATCAGCTGAAGATCA 104
```


